This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of
the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS

- TEXT CUT OFF AT TOP, BOTTOM OR SIDES

- FADED TEXT

- ILLEGIBLE TEXT

- SKEWED/SLANTED IMAGES

- COLORED PHOTOS

- BLACK OR VERY BLACK AND WHITE DARK PHOTOS

- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

(54) Title: IMPROVED EPITODE DISPLAYING PHAGE

(57) Abstract

The present invention relates to libraries of display phage which display various mutant epitopic peptides or protein domains with potential to bind to a target material of interest. Site-specific protease cleavable linkers may be incorporated into these phages to overcome the problem of irreversible binding to target materials. When the displayed insert is a mini-protein, a reagent which cleaves the crosslink may be used to facilitate release of a tightly bound display phage. The invention also relates to providing a more balanced library of epitopic peptide display phage.

1

IMPROVED EPITOPE DISPLAYING PHAGE

BACKGROUND OF THE INVENTION

Field of the Invention

5   This invention relates to the expression and display
of libraries of mutated epitopic peptides or potential
binding protein domains on the surface of phage, and the
screening of those libraries to identify high affinity
species.

10 ### Information Disclosure Statement

The amino acid sequence of a protein determines its
three-dimensional (3D) structure, which in turn
determines protein function. Some residues on the
polypeptide chain are more important than others in
15 determining the 3D structure of a protein. Substitutions
of amino acids that are exposed to solvent are less
likely to affect the 3D structure than are substitutions
at internal loci.

"Protein engineering" is the art of manipulating the
20 sequence of a protein in order to alter its binding
characteristics. The factors affecting protein binding
are known, but designing new complementary surfaces has
proven difficult.

With the development of recombinant DNA techniques,
25 it became possible to obtain a mutant protein by mutating
the gene encoding the native protein and then expressing
the mutated gene. Several mutagenesis strategies are
known. One, "protein surgery", involves the introduction
of one or more predetermined mutations within the gene of
30 choice. A single polypeptide of completely predetermined
sequence is expressed, and its binding characteristics
ar evaluated.

At the other extreme is random mutagenesis by means
of relatively nonspecific mutagens such as radiation and
35 various chemical agents.

2

It is possible to randomly vary predetermined
nucleotides using a mixture of bases in the appropriate
cycles of a nucleic acid synthesis procedure. The
proportion of bases in the mixture, for each position of
5   a codon, will determine the frequency at which each amino
acid will occur in the polypeptides expressed from the
degenerate DNA population. Oliphant et al. (OLIP86) and
Oliphant and Struhl (OLIP87) have demonstrated ligation
and cloning of highly degenerate oligonucleotides, which
10   were used in the mutation of promoters. They suggested
that similar methods could be used in the variation of
protein coding regions. They do not say how one should:
a) choose protein residues to vary, or b) select or
screen mutants with desirable properties. Reidhaar-Olson
15   and Sauer (REID88a) have used synthetic degenerate oligo-
nts to vary simultaneously two or three residues through
all twenty amino acids. See also Vershon et al.
(VERS86a; VERS86b). Reidhaar-Olson and Sauer do not
discuss the limits on how many residues could be varied
20   at once nor do they mention the problem of unequal
abundance of DNA encoding different amino acids.

A number of researchers have directed unmutated
foreign antigenic epitopes to the surface of phage, fused
to a native phage surface protein, and demonstrated that
25   the epitopes were recognized by antibodies.

Dulbecco (DULB86) suggests a procedure for incor-
porating a foreign antigenic epitope into a viral surface
protein so that the expressed chimeric protein is dis-
played on the surface of the virus in a manner such that
30   the foreign epitope is accessible to antibody. In 1985
Smith (SMIT85) reported inserting a nonfunctional segment
of the EcoRI endonuclease gene into gene III of bacterio-
phage f1, "in phase". The gene III protein is a minor
coat protein necessary for infectivity.   Smith demons-
35   trated that the recombinant phage were adsorbed by
immobilized antibody raised against the EcoRI endonucle-

3

ase, and could be eluted with acid. De la Cruz et al.
(DELA88) have expressed a fragment of the repeat region
of the circumsporozoite protein from _Plasmodium_
falciparum on the surface of M13 as an insert in the gene
5   III protein. They showed that the recombinant phage were
both antigenic and immunogenic in rabbits, and that such
recombinant phage could be used for B epitope mapping.
The researchers suggest that similar recombinant phage
could be used for T epitope mapping and for vaccine
10   development.

McCafferty et al. (MCCA90) expressed a fusion of an
Fv fragment of an antibody to the N-terminal of the pIII
protein. The Fv fragment was not mutated.

Ladner, Glick, and Bird, WO88/06630 (publ. 7 Sept.
15   1988 and having priority from US application 07/021,046,
assigned to Genex Corp.) (LGB) speculate that diverse
single chain antibody domains (SCAD) may be screened for
binding to a particular antigen by varying the DNA
encoding the combining determining regions of a single
20   chain antibody, subcloning the SCAD gene into the gpV
gene of phage lambda so that a SCAD/gpV chimera is
displayed on the outer surface of phage lambda, and
selecting phage which bind to the antigen through
affinity chromatography.

25   Parmley and Smith (PARM88) suggested that an epitope
library that exhibits all possible hexapeptides could be
constructed and used to isolate epitopes that bind to
antibodies.   In discussing the epitope library, the
authors did not suggest that it was desirable to balance
30   the representation of different amino acids. Nor did
they teach that the insert should encode a complete
domain of the exogenous protein. Epitopes are considered
to be unstructured peptides as opposed to structured
proteins. Scott and Smith (SCOT90) and Cwirla et al.
35   (CWIR90) prepared "epitope libraries" in which potential
hexapeptide epitopes for a target antibody were randomly

4

mutated by fusing degenerate oligonucleotides, encoding
the epitopes, with gene III of fd phage, and expressing
the fused gene in phage-infected cells. The cells
manufactured fusion phage which displayed the epitopes on
5   their surface; the phage which bound to immobilized
antibody were eluted with acid and studied. Devlin et
al. (DEVL90) similarly screened, using M13 phage, for
random 15 residue epitopes recognized by streptavidin.

The Scott and Smith, Cwirla et al., and Devlin et
10   al., libraries provided a highly biased sampling of the
possible amino acids at each position. Their primary
concern in designing the degenerate oligonucleotide
encoding their variable region was to ensure that all
twenty amino acids were encodible at each position; a
15   secondary consideration was minimizing the frequency of
occurrence of stop signals. Consequently, Scott and
Smith and Cwirla et al. employed NNK (N=equal mixture of
G, A, T, C; K=equal mixture of G and T) while Devlin et
al. used NNS (S=equal mixture of G and C). There was no
20   attempt to minimize the frequency ratio of most favored-
to-least favored amino acid, or to equalize the rate of
occurrence of acidic and basic amino acids.

Devlin et al. characterized several affinity-
selected streptavidin-binding peptides, but did not
25   measure the affinity constants for these peptides.
Cwirla et al. did determine the affinity constant for his
peptides, but were disappointed to find that his best
hexapeptides had affinities (350-300nM), "orders of
magnitude" weaker than that of the native Met-enkephalin
30   epitope (7nM) recognized by the target antibody. Cwirla
et al. speculated that phage bearing peptides with higher
affinities remained bound under acidic elution, possibly
because of multivalent interactions between phage (carry-
ing about 4 copies of pIII) and the divalent target IgG.
35   Scott and Smith were able to find peptides whose affinity
for the target antibody (A2) was comparable to that of

5

the reference myohemerythrin epitope (50nM). However,
Scott and Smith likewise expressed concern that some
high-affinity peptides were lost, possibly through
irreversible binding of fusion phage to target.

5   Ladner, et al, WO90/02809, incorporated by reference
herein, describe a process for the generation and
identification of novel binding proteins having affinity
for a predetermined target. In this process, a gene
encoding a potential binding domain (as distinct from a
10   mere epitopic peptide), said gene being obtained by
random mutagenesis of a limited number of predetermined
codons, is fused to a genetic element which causes the
resulting chimeric expression product to be displayed on
the outer surface of a virus (especially a filamentous
15   phage) or a cell. Chromatographic selection is then used
to identify viruses or cells whose genome includes such
a fused gene which coded for the protein which bound to
the chromatographic target. Ladner, et al. discusss
several methods of recovering the gene of interest when
20   the viruses or cells is so tightly bound to the target
that it cannot be washed off in viable form. These are
growing them in situ on the chromatographic matrix,
fragmenting the matrix and using it as an inoculant into
a culture vessel, degrading the linkage between the
25   matrix and the target material, and degrading the viruses
or cells but then recovering their DNA. However, these
methods will also recover viruses or cells which are
nonspecifically bound to the target material.
          WO90/02809 also addressed strategies for
30   mutagenesis, including one which provides all twenty
amino acids in substantially equal proportions, but only
in the context of mutagenesis of protein domains, not
epitopic peptides.

35   SUMMARY OF THE INVENTION

6

      The present invention is intended to overcome the deficiencies discussed above. In one embodiment of the invention, a library of "display phage" is used to identify binding domains with a high affinity for a

5   predetermined target. Potential binding domains are displayed on the surface of the phage. This is achieved by expressing a fused gene which encodes a chimeric outer surface protein comprising the potential binding domain and at least a functional portion of a coat protein

10   native to the phage. The preferred embodiment uses a pattern of semirandom mutagenesis, called "variegation", that focuses mutations into those residues of a parental binding domain that are most likely to affect its binding properties and are least likely to destroy its underlying

15   structure. As a result, while any one phage displays only a single foreign binding domain (though possibly in multiple copies), the phage library collectively displays thousands, even millions, of different binding domains. The phage library is screened by affinity separation

20   techniques to identify those phage bearing successful (high affinity) binding domains, and these phage are recovered and characterized to determine the sequence of the successful binding domains. These successful binding domains may then serve as the parental binding domains

25   for another round of variegation and affinity separation.

      In another embodiment of the invention, the display phage display on their surface a chimeric outer surface protein comprising a functional portion of a native outer surface protein and a potential epitope. In an epitope

30   library made of these display phage, the region corresponding to the foreign epitope is hypervariable. The library is screened with an antibody or other binding protein of int rest and high affinity epitopes are identified. References to display, mutagenesis and

35   screening of potential binding domains should be taken to

7

apply, _mutatis mutandis_, to display, mutagenesis and
screening of potential epitopes, unless stated otherwise.

    As previously mentioned, when several copies of the
chimeric coat protein are displayed on a single phage,
5   there is a risk that irreversible binding will occur,
especially if the target is multivalent (as with an
antibody). In this case, the phage last eluted by an
elution gradient will not be the ones bearing the highest
affinity epitopes or binding domains, but rather will be
10   those having an affinity high enough to hold on to the
target under the initial elution conditions but not so
high as to bind irreversibly. As a result, the
methodology known in the art may fail to recover very
high affinity epitopes or binding domains, which for many
15   purposes are the most desirable species.

    We propose to cope with the problem of irreversible
binding by incorporating into the chimeric coat protein
a linker sequence, between the foreign epitope or binding
domain, and the sequence native to the wild-type phage
20   coat protein, which is cleavable by a site-specific
protease. In this case, the phage library is incubated
with the immobilized target. Lower affinity phage are
eluted off the target and only the solid phase (bearing
the high affinity phage) is retained. The aforementioned
25   linker sequence is cleaved, and the phage particles are
released, leaving the bound epitope or binding domain
behind. One may then recover the particles (and sequence
their DNA to determine the sequence of the corresponding
epitope or binding domain) or the bound peptide (and
30   sequence its amino acids directly). The former recovery
method is preferred, as the encoding DNA may be amplified
in vitro using PCR or in vivo by transfecting suitable
host cells with the high affinity display phage. While
the production of fusion proteins with cleavable linkers
35   is known in the art, the use of such linkers to

8

facilitate controlled cleavage of a chimeric coat protein
of a phage has not previously been reported.

   Another method of addressing irreversible binding is
appropriate when the binding domains are "mini-proteins",
5   i.e., relatively small peptides whose stability of
structure primarily attributable to the presence of one
or more covalent crosslinks, e.g., disulfide bonds. As
in the example above, low affinity phage are removed
first. The remaining, high affinity, bound phage are
10   then treated with a reagent which breaks the crosslink,
such as dithiothreitol in the case of a domain with
disulfide bonds, but does not cleave peptidyl bonds or
modify the side chains of amino acids which are not
crosslinked. This will usually result in sufficient
15   denaturation to either release the phage outright or to
permit their elution by other means.

   These two methods, of course, are not mutually
exclusive.

   In the previously known epitope display phage
20   libraries, the phage genome was altered by replacing the
gene encoding the wild-type gene III protein of M13 with
one encoding a chimeric coat protein. As a result, the
five normal copies of the wild type gene III protein were
all replaced by the chimeric coat protein, whereby each
25   phage had five potential binding sites for the target,
and hence a very high potential avidity. With high
affinity epitopes (or binding domains), this might well
contribute to irreversible binding.

   One method of the present invention of reducing the
30   avidity of display phage, especially epitope display
phage, for their target, and hence of alleviating the
problem of irreversible binding, is to engineer the phage
to contain two genes that each express a coat protein,
one encoding the wild type coat protein, and the other
35   the cognate chimeric coat protein. Thus, phage bearing
identical epitopes or binding domains may yet bear

9

different ratios of wild type to chimeric coat protein
molecules, and hence have different avidities.   The
average ratio for the library will be dependent on the
relative levels of expression of the two cognate genes.

5   It may be advantageous to be able to modulate the
ratio of the chimeric coat protein to its cognate wild-
type   coat   protein.   For   example,   early   in   the
evolutionary process, the affinity of the binding domains
for their target may be rather low, especially if they

10   are based on a parental binding domain which has no
affinity for the target.

Modulation may be achieved by placing the chimeric
gene under the control of a regulatable promoter.

While it may be possible to place the cognate wild-

15   type gene under the control of a second, differently
regulated promoter, this may be impracticible if, as with
the M13 geneIII, the gene is part of a polycistronic
operon.   In this case, expression of the wild-type gene
may be reduced by replacing its methionine initiation

20   codon with a leucine initiation codon.

M13 gene III, as previously noted, encodes one of
the minor coat proteins of this filamentous phage (five
copies per phage).   In view of the difficulties with
irreversible binding reported by those modifying this

25   gene so that a foreign epitope is displayed on the phage
coat, use of the M13 major coat protein was clearly
discouraged.   However, we have found that chimeric major
coat proteins are in fact useful for displaying potential
binding domains for screening purposes even though

30   (indeed, sometimes because) there are over a thousand
copies of this protein per phage.   It is believed that
the major (VIII) coat protein would likewise be useful in
constructing an epitope phage library.

We have also developed a linker suitable for

35   attaching potential binding domains (or epitopic

peptides) to this major coat protein, and perhaps to other proteins as well.

Finally, to the extent that some of the problems experienced with epitope libraries have been attributable
5   to the use of patterns of mutagenesis which lead to highly biased allocations of amino acids, the present invention is also directed to a variety of improved patterns that lead to less biased and hence more efficient epitope phage libraries.
10

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows how a phage may be used as a genetic phage. At (a) we have a wild-type precoat protein lodged in the lipid bilayer. The signal peptide is in the
15   periplasmic space. At (b), a chimeric precoat protein, with a potential binding domain interposed between the signal peptide and the mature coat protein sequence, is similarly trapped. At (c) and (d), the signal peptide has been cleaved off the wild-type and chimeric proteins,
20   respectively, but certain residues of the coat protein sequence interact with the lipid bilayer to prevent the mature protein from passing entirely into the periplasm. At (e) and (f), mature wild-type and chimeric protein are assembled into the coat of a single stranded DNA phage as
25   it emerges into the periplasmic space. The phage will pass through the outer membrane into the medium where it can be recovered and chromatographically evaluated.

Figure 2 shows the $C_\alpha s$ of the coat protein of phage f1.

30   ## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

I. DISPLAY STRATEGY

A. General Considerations

The present invention contemplates that a potential binding domain (pbd) or a potential epitope will be dis-
35   played on the surface of a phage in the form of a fusion with a coat (outer surface) protein (OSP) of the phage.

11

This chimeric outer surface protein is the processed
product of the polypeptide expressed by an display gene
inserted into the phage genome; therefore: 1) the genome
of the phage must allow introduction of the display gene
5   either by tolerating additional genetic material or by
having replaceable genetic material; 2) the virion must
be capable of packaging the genome after accepting the
insertion or substitution of genetic material, and 3) the
display of the OSP-IPBD protein on the phage surface must
10   not disrupt virion structure sufficiently to interfere
with phage propagation.

When the viral particle is assembled, its coat
proteins may attach themselves to the phage: a) from the
cytoplasm, b) from the periplasm, or c) from within the
15   lipid bilayer. The immediate expression product of the
display gene must feature, at its amino terminal, a
functional secretion signal peptide, such as the phoA
signal (MKQSTIALALLPLLFTPVTKA), if the coat protein
attaches to the phage from the periplasm or from within
20   the lipid bilayer. If a secretion signal is necessary
for the display of the potential binding domain, in an
especially preferred embodiment the bacterial cell in
which the hybrid gene is expressed is of a "secretion-
permissive" strain.

25   The DNA sequence encoding the foreign epitope or
binding domain should precede the sequence encoding the
coat protein proper if the amino terminal of the
processed coat protein is normally its free end, and
should follow it if the carboxy terminal is the normal
30   free end.

The morphogenetic pathway of the phage determines
the environment in which the IPBD will have opportunity
to fold. Periplasmically assembled phage are preferred
when IPBDs contain essential disulfides, as such IPBDs
35   may not fold within a cell (these proteins may fold after
the phage is released from the cell). Intracellularly

12

assembled phage are preferred when the IPBD needs large
or insoluble prosthetic groups (such as Fe_4S_4 clusters),
since the IPBD may not fold if secreted because the
prosthetic group is lacking.

5   When variegation is introduced, multiple infections
could generate hybrid GPs that carry the gene for one PBD
but have at least some copies of a different PBD on their
surfaces; it is preferable to minimize this possibility
by infecting cells with phage under conditions resulting

10   in a low multiple-of-infection (MOI).

       For a given bacteriophage, the preferred OSP is
usually one that is present on the phage surface in the
largest number of copies, as this allows the greatest
flexibility in varying the ratio of OSP-IPBD to wild type

15   OSP and also gives the highest likelihood of obtaining
satisfactory affinity separation. Moreover, a protein
present in only one or a few copies usually performs an
essential function in morphogenesis or infection;
mutating such a protein by addition or insertion is

20   likely to result in reduction in viability of the GP.
Nevertheless, an OSP such as M13 gIII protein may be an
excellent choice as OSP to cause display of the PBD.

       It is preferred that the wild-type osp gene be pre-
served.   The ipbd gene fragment may be inserted either

25   into a second copy of the recipient osp gene or into a
novel engineered osp gene. It is preferred that the osp-
ipbd gene be placed under control of a regulated
promoter. Our process forces the evolution of the PBDs
derived from IPBD so that some of them develop a novel

30   function, viz. binding to a chosen target. Placing the
gene that is subject to evolution on a duplicate gene is
an imitation of the widely-accepted scenario for the
evolution of protein families.   It is now generally
accepted that gene duplication is the first step in the

35   evolution of a protein family from an ancestral protein.
By having two copies of a gene, the affected

physiological process can tolerate mutations in one of
the genes.   This process is well understood and
documented for the globin family (cf. DICK83, p65ff, and
CREI84, p117-125).

5   The user must choose a site in the candidate OSP
gene for inserting a ipbd gene fragment.   The coats of
most bacteriophage are highly ordered. Filamentous phage
can be described by a helical lattice; isometric phage,
by an icosahedral lattice.   Each monomer of each major
10   coat protein sits on a lattice point and makes defined
interactions with each of its neighbors. Proteins that
fit into the lattice by making some, but not all, of the
normal lattice contacts are likely to destabilize the
virion by: a) aborting formation of the virion, b) making
15   the virion unstable, or c) leaving gaps in the virion so
that the nucleic acid is not protected.   Thus in
bacteriophage, it is important to retain in engineered
OSP-IPBD fusion proteins those residues of the parental
OSP that interact with other proteins in the virion. For
20   M13 gVIII, we prefer to retain the entire mature protein,
while for M13 gIII, it might suffice to retain the last
100 residues (or even fewer).   Such a truncated gIII
protein would be expressed in parallel with the complete
gIII protein, as gIII protein is required for phage
25   infectivity.
         The display gene is placed downstream of a known
promoter, preferably a regulated promoter such as lacUV5,
tac, or trp.
         B. Filamentous Phages
30   The filamentous phages, which include M13, f1, fd,
If1, Ike, Xf, Pf1, and Pf3, are of particular interest.
The entire life cycle of the filamentous phage M13, a
c mmon cloning and sequencing vector, is well understood.
The genetic structure (SCHA78) of M13 is well known as is
35   the physical structure of the virion (BANN81, BOEK80,
CHAN79, ITOK79, KAPL78, KUHN85b, KUHN87, MAKO80, MARV78,

14

MESS78, OHKA81, RASC86, RUSS81, SCHA78, SMIT85, WEBS78,
and ZIMM82); see RASC86 for a recent review of the
structure and function of the coat proteins.

    Marvin and collaborators (MARV78, MAKO80, BANN81)
5   have determined an approximate 3D virion structure of the
closely related phage f1 by a combination of genetics,
biochemistry, and X-ray diffraction from fibers of the
virus. Figure 2 is drawn after the model of Banner et
$al.$ (BANN81) and shows only the C_{α}s of the protein. The
10   apparent holes in the cylindrical sheath are actually
filled by protein side groups so that the DNA within is
protected. The amino terminus of each protein monomer is
to the outside of the cylinder, while the carboxy
terminus is at smaller radius, near the DNA. Although
15   other filamentous phages (e.g. Pf1 or Ike) have different
helical symmetry, all have coats composed of many short
α-helical monomers with the amino terminus of each
monomer on the virion surface.

    1. M13 Major Coat Protein (gVIII)
20   The major coat protein of M13 is encoded by gene
VIII. The 50 amino acid mature gene VIII coat protein is
synthesized as a 73 amino acid precoat (SCHA78; ITOK79).
The first 23 amino acids constitute a typical signal-
sequence which causes the nascent polypeptide to be
25   inserted into the inner cell membrane. Whether the
precoat inserts into the membrane by itself or through
the action of host secretion components, such as SecA and
SecY, remains controversial, but has no effect on the
operation of the present invention.

30   An $\underline{E. \; coli}$ signal peptidase (SP-I) recognizes amino
acids 18, 21, and 23, and, to a lesser extent, residue
22, and cuts between residues 23 and 24 of the precoat
(KUHN85a, KUHN85b, OLIV87). After removal of the signal
sequence, the amino terminus of the mature coat is
35   located on the periplasmic side of the inner membrane;
the carboxy terminus is on the cytoplasmic side. About

15

3000 copies of the mature 50 amino acid coat protein associate side-by-side in the inner membrane.

We have constructed a tripartite gene comprising:

1) DNA encoding a signal sequence directing secretion
5   of parts (2) and (3) through the inner membrane,

2) DNA encoding the mature BPTI sequence, and

3) DNA encoding the mature M13 gVIII protein.

This gene causes BPTI to appear in active form on the surface of M13 phage.

10   2. M13 Minor Coat Proteins, Generally

An introduced binding domain or epitope may also be displayed on a filamentous phage as a portion of a chimeric minor coat protein. These are encoded by genes III, VI, VII, and IX, and each is present in about 5
15 copies per virion and is related to morphogenesis or infection. In contrast, the major coat protein is present in more than 2500 copies per virion. The gene III, VI, VII, and IX proteins are present at the ends of the virion.

20   3. The M13 gIII Minor Coat Protein

The single-stranded circular phage DNA associates with about five copies of the gene III protein and is then extruded through the patch of membrane-associated coat protein in such a way that the DNA is encased in a
25 helical sheath of protein (WEBS78). The DNA does not base pair (that would impose severe restrictions on the virus genome); rather the bases intercalate with each other independent of sequence.

Smith (SMIT85) and de la Cruz et al. (DELA88) have
30 shown that insertions into gene III cause novel protein domains to appear on the virion outer surface. The mini-protein's gene may be fused to gene III at the site used by Smith and by de la Cruz et al., at a codon correspond-ing to another domain boundary or to a surface loop of
35 the protein, or to the amino terminus of the mature protein.

All published works use a vector containing a single modified gene III of fd. Thus, all five copies of gIII are identically modified. Gene III is quite large (1272 b.p. or about 20% of the phage genome) and it is
5 uncertain whether a duplicate of the whole gene can be stably inserted into the phage. Furthermore, all five copies of gIII protein are at one end of the virion. When bivalent target molecules (such as antibodies) bind a pentavalent phage, the resulting complex may be
10 irreversible. Irreversible binding of the phage to the target greatly interferes with affinity enrichment of the GPs that carry the genetic sequences encoding the novel polypeptide having the highest affinity for the target.
    To reduce the likelihood of formation of
15 irreversible complexes, we may use a second, synthetic gene that encodes only the carboxy-terminal portion of III. We might, for example, engineer a gene that comprises (from 5' to 3'):
    1) a promoter (preferably regulated),
20 2) a ribosome-binding site,
    3) an initiation codon,
    4) a sequence encoding a functional signal peptide directing secretion of parts (5) and (6) through the inner membrane,
25 5) DNA encoding a potential binding domain,
    6) DNA encoding residues 275 through 424 of M13 gIII protein,
    7) a translation stop codon, and
    8) (optionally) a transcription stop signal.
30 Note that in the gIII protein, the amino terminal moiety is responsible for pilus binding (i.e., for infectivity) and the carboxy terminal moiety for packaging, so that the chimeric gIII protein described above is able to assemble into the viral coat, but does
35 not contribute to infectivity.

17

We leave the wild-type gene _III_ so that some
unaltered gene III protein will be present.

Thus, the hybrid gene may comprise DNA encoding a
potential binding domain operably linked to a signal
5   sequence (_e.g._, the signal sequences of the bacterial
phoA or _bla_ genes or the signal sequence of M13 phage
geneIII)   and to DNA encoding at least a functional
portion of a coat protein (_e.g._, the M13 gene III or gene
VIII proteins) of a filamentous phage (_e.g._, M13). The
10   expression product is transported to the inner membrane
(lipid bilayer) of the host cell, whereupon the signal
peptide is cleaved off to leave a processed hybrid
protein. The C-terminus of the coat protein-like com-
ponent of this hybrid protein is trapped in the lipid
15   bilayer, so that the hybrid protein does not escape into
the periplasmic space. (This is typical of the wild-type
coat protein.) As the single-stranded DNA of the nascent
phage particle passes into the periplasmic space, it
collects both wild-type coat protein and the hybrid
20   protein from the lipid bilayer. The hybrid protein is
thus phaged into the surface sheath of the filamentous
phage, leaving the potential binding domain exposed on
its outer surface.

4. Coat Proteins of Pf3

25   Similar constructions could be made with other
filamentous phage. Pf3 is a well known filamentous phage
that infects _Pseudomonas aerugenosa_ cells that harbor an
IncP-1 plasmid. The entire genome has been sequenced
(LUIT85) and the genetic signals involved in replication
30   and assembly are known (LUIT87). The major coat protein
of PF3 is unusual in having no signal peptide to direct
its secretion. The sequence has charged residues ASP_7,
ARG_{37}, LYS_{40}, and PHE_{44}-COO⁻ which is consistent with the
amino terminus being exposed. Thus, to cause an IPBD to
35   appear on the surface of Pf3, we construct a tripartite
gene comprising:

18

    1) a signal sequence known to cause secretion in P.
       aerugenosa (preferably known to cause secretion of
       IPBD) fused in-frame to,

    2) a gene fragment encoding the IPBD sequence, fused
5   in-frame to,

    3) DNA encoding the mature Pf3 coat protein.

Optionally, DNA encoding a flexible linker of one to 10
amino acids is introduced between the ipbd gene fragment
and the Pf3 coat-protein gene. Optionally, DNA encoding
10 the recognition site for a specific protease, such as
tissue plasminogen activator or blood clotting Factor Xa,
is introduced between the ipbd gene fragment and the Pf3
coat-protein gene. Amino acids that form the recognition
site for a specific protease may also serve the function
15 of a flexible linker. This tripartite gene is introduced
into Pf3 so that it does not interfere with expression of
any Pf3 genes. To reduce the possibility of genetic
recombination, part (3) is designed to have numerous
silent mutations relative to the wild-type gene. Once
20 the signal sequence is cleaved off, the IPBD is in the
periplasm and the mature coat protein acts as an anchor
and phage-assembly signal. It does not matter that this
fusion protein comes to rest in the lipid bilayer by a
route different from the route followed by the wild-type
25 coat protein.

Gene Construction

    The structural coding sequence of the display gene
encodes a chimeric coat protein and any required
secretion signal. A "chimeric coat protein" is a fusion
30 of a first amino acid sequence (essentially corresponding
to at least a functional portion of a phage coat protein)
with a second amino acid sequence, e.g., a domain foreign
to and not substantially homologous with any domain of
the first protein. A chimeric protein may present a
35 foreign domain which is found (albeit in a different
protein) in an organism which also expresses the first

19

protein, or it may be an "interspecies", "intergeneric",
etc. fusion of protein structures expressed by different
kinds of organisms. The foreign domain may appear at the
amino or carboxy terminal of the first amino acid
5   sequence (with or without an intervening spacer), or it
may interrupt the first amino acid sequence. The first
amino acid sequence may correspond exactly to a surface
protein of the phage, or it may be modified, e.g., to
facilitate the display of the binding domain.

10   A preferred site for insertion of the ipbd gene into
the phage osp gene is one in which: a) the IPBD folds
into its original shape, b) the OSP domains fold into
their original shapes, and c) there is no interference
between the two domains.

15   If there is a model of the phage that indicates that
either the amino or carboxy terminus of an OSP is exposed
to solvent, then the exposed terminus of that mature OSP
becomes the prime candidate for insertion of the ipbd
gene. A low resolution 3D model suffices.

20   In the absence of a 3D structure, the amino and
carboxy termini of the mature OSP are the best candidates
for insertion of the ipbd gene. A functional fusion may
require additional residues between the IPBD and OSP
domains to avoid unwanted interactions between the
25   domains. Random-sequence DNA or DNA coding for a
specific sequence of a protein homologous to the IPBD or
OSP, can be inserted between the osp fragment and the
ipbd fragment if needed.

        Fusion at a domain boundary within the OSP is also
30   a good approach for obtaining a functional fusion. Smith
exploited such a boundary when subcloning heterologous
DNA into gene III of f1 (SMIT85).

        The criteria for identifying OSP domains suitable
for causing display of an IPBD are somewhat different
35   from those used to identify and IPBD. When identifying
an OSP, minimal size is not so important because the OSP

20

domain will not appear in the final binding molecule nor
will we need to synthesize the gene repeatedly in each
variegation round. The major design concerns are that:
a) the OSP::IPBD fusion causes display of IPBD, b) the
5   initial genetic construction be reasonably convenient,
and c) the osp::ipbd gene be genetically stable and
easily manipulated.   There are several methods of
identifying domains.   Methods that rely on atomic
coordinates have been reviewed by Janin and Chothia
10 (JANI85).   These methods use matrices of distances
between α carbons (C$_\alpha$), dividing planes (cf. ROSE85), or
buried surface (RASH84). Chothia and collaborators have
correlated the behavior of many natural proteins with
domain structure (according to their definition). Rashin
15 correctly predicted the stability of a domain comprising
residues 206-316 of thermolysin (VITA84, RASH84).

Many researchers have used partial proteolysis and
protein sequence analysis to isolate and identify stable
domains. (See, for example, VITA84, POTE83, SCOT87a, and
20 PABO79.)   Pabo et al. used calorimetry as an indicator
that the cI repressor from the coliphage)\ contains two
domains; they then used partial proteolysis to determine
the location of the domain boundary.

If the only structural information available is the
25 amino acid sequence of the candidate OSP, we can use the
sequence to predict turns and loops.   There is a high
probability that some of the loops and turns will be
correctly predicted (cf. Chou and Fasman, (CHOU74));
these locations are also candidates for insertion of the
30 ipbd gene fragment.

Fusing one or more new domains to a protein may make
the ability of the new protein to be exported from the
cell different from the ability of the parental protein.
The signal peptide of the wild-type coat protein may
35 function for authentic polypeptide but be unable to
direct export of a fusion. To utilize the Sec-dependent

21

pathway, one may need a different signal peptide. Thus,
to express and display a chimeric BPTI/M13 gene VIII
protein, we found it necessary to utilize a heterologous
signal peptide (that of phoA).

5   Phage that display peptides having high affinity for
the target may be quite difficult to elute from the
target, particularly a multivalent target. One can
introduce a cleavage site for a specific protease, such
as blood-clotting Factor Xa, into the fusion protein so
10   that the binding domain can be cleaved from the genetic
package. Such cleavage has the advantage that all
resulting phage have identical coat proteins and
therefore are equally infective, even if polypeptide-
displaying phage can be eluted from the affinity matrix
15   without cleavage. This step allows recovery of valuable
genes which might otherwise be lost. To our knowledge,
no one has disclosed or suggested using a specific
protease as a means to recover an information-containing
genetic package or of converting a population of phage
20   that vary in infectivity into phage having identical
infectivity.

There exist a number of highly specific proteases.
While the invention does not reside in the choice of any
particular protease, the protease is preferably
25   sufficiently specific so that under the cleavage
conditions, it will cleave the linker but not any
polypeptide essential to the viability of the phage, or
.(save in rare cases) the potential epitope/binding
domain. It is possible that choice of particular
30   cleavage conditions, e.g., low temperature, may make it
feasible to use a protease that would otherwise be
unsuitable.

The blood-clotting and complementation systems
contains a number of very specific proteases. Usually,
35   the enzymes at the early stages of a cascade are more
specific than are the later ones. For example, Factor X_a

(F.X_a) is more specific than is thrombin (cp. Table 10-2
of COLM87). Bovine F.X_a cleaves after the sequence Ile-
Glu-Gly-Arg while human F.X_a cleaves after Ile-Asp-Gly-
Arg.   Either protease-linker pair may be used, as
5   desired.   If thrombin is used, the most preferred
thrombin-sensitive linkers are those found in fibrinogen,
Factor XIII, and prothrombin. Preferably, one would take
the linker sequence from the species from which the
thrombin is obtained; for example, if bovine thrombin is
10   to be used, then one uses a linker taken from bovine
fibrinogen or bovine F.XIII.
         Human Factor XI_a cleaves human Factor IX at two
places (COLM87, p.42):
         Q T $\underline{S\ K\ L\ T\ R_{145}}$/ $\underline{A}$ E A V F   and
15   S F $\underline{N\ D\ F\ T\ R_{180}}$/ $\underline{V}$ V G G E.
Thus one could incorporate either of these sequences
(especially the underscored portions) as a linker between
the PBD and the GP-surface-anchor domain (GPSAD) and use
human F.XI_a to cleave them.
20   Human kallikrein cuts human F.XII at R_{353} (COLM87,
p.258):
         L F $\underline{S\ S\ M\ T\ R_{353}}$/ $\underline{V\ V\ G}$ G L V.
This sequence has significant similarity to the hF.XI_a
sites above. One could incorporate the sequence SSMTRVVG
25   as a linker between PBD and GPSAD and cleave PBD from the
GP with human kallikrein.
         Human F.XII_a cuts human F.XI at R_{369} (COLM87, p.256):
         $\underline{K\ I\ K\ P\ R_{369}}$/ $\underline{I\ V\ G}$ G T.
One could incorporate KIKPRIVG as a linker between PBD
30   and GPSAD.   PBD could then be cleaved from GP with
hF.XII_a.
         Other proteases that have been used to cleave fusion
proteins include enterokinas , collagenase, chymosin,
urokinase, renin, and certain signal peptidases.   See
35   Rutter, US 4,769,326.

23

When a protease inhibitor is sought, the target
protease and other proteases having similar substrate
specificity are not preferred for cleaving the PBD from
the GP.   It is preferred that a linker resembling the
5   substrate of the target protease not be incorporated
anywhere on the display phage because this could make
separation of excellent binders from the rest of the
population needlessly more difficult.

If there is steric hindrance of the site-specific
10   cleavage of the linker, the linker may be modified so
that the cleavage site is more exposed, e.g., by
interposing glycines (for additional flexibility) or
prolines (for maximum elongation) between the cleavage
site and the bulk of the protein.   GUAN91 improved
15   thrombin cleavage of a GST fusion protein by introducing
a glycine-rich linker (PGISGGGGG) immediately after the
thrombin cleavage site (LVPRGS).   A suitable linker may
also be identified by variegation-and-selection.

The sequences of regulatory parts of the gene are
20   taken from the sequences of natural regulatory elements:
a) promoters, b) Shine-Dalgarno sequences, and c) trans-
criptional terminators.   Regulatory elements could also
be designed from knowledge of consensus sequences of
natural regulatory regions.   The sequences of these
25   regulatory elements are connected to the coding regions;
restriction sites are also inserted in or adjacent to the
regulatory regions to allow convenient manipulation.

The essential function of the affinity separation is
to separate GPs that bear PBDs (derived from IPBD) having
30   high affinity for the target from GPs bearing PBDs having
low affinity for the target.   If the elution volume of a
phage depends on the number of PBDs on the phage surface,
then a phage bearing many PBDs with low affinity,
$GP(PBD_w)$, might co-elute with a phage bearing fewer PBDs
35   with high affinity, $GP(PBD_s)$.   Regulation of the display
gene preferably is such that most phage display

24

sufficient PBD to effect a good separation according to
affinity. Use of a regulatable promoter to control the
level of expression of the display gene allows fine
adjustment of the chromatographic behavior of the
5   variegated population.

Induction of synthesis of engineered genes in
vegetative bacterial cells has been exercised through the
use of regulated promoters such as lacUV5, trpP, or tac
(MANI82). The factors that regulate the quantity of
10   protein synthesized are sufficiently well understood that
a wide variety of heterologous proteins can now be
produced in E. coli, B. subtilis and other host cells in
at least moderate quantities (SKER88, BETT88).
Preferably, the promoter for the display gene is subject
15   to regulation by a small chemical inducer. For example,
the lac promoter and the hybrid trp-lac (tac) promoter
are regulatable with isopropyl thiogalactoside (IPTG).
The promoter for the constructed gene need not come from
a natural osp gene; any regulatable promoter functional
20   in bacteria can be used. A non-leaky promoter is
preferred.

The coding portions of genes to be synthesized are
designed at the protein level and then encoded in DNA.
While the primary consideration in devising the DNA
25   sequence is obtaining the desired diverse population of
potential binding domains (or epitopes), consideration is
also given to providing restriction sites to facilitate
further gene manipulation, minimizing the potential for
recombination and spontaneous mutation, and achieving
30   efficient translation in the chosen host cells.

The present invention is not limited to any parti-
cular method of DNA synthesis or construction. Conven-
tional DNA synthesizers may be used, with appropriate
reagent modifications for production of variegated DNA
35   (similar to that now used for production of mixed
probes).

25

The phage are genetically engineered and then transfected into host cells, e.g., E. coli, B. subtilis, or P. aeruginosa, suitable for amplification. The present invention is not limited to any one method of
5   transforming cells with DNA or to any particular host cells.

THE INITIAL POTENTIAL BINDING DOMAIN (IPBD):

By virtue of the present invention, proteins may be obtained which can bind specifically to targets other
10   than the antigen-combining sites of antibodies. For the purposes of the appended claims, a protein P is a binding protein if for some target other than the variable domain of an antibody, the dissociation constant K_D (P,A) < 10^{-6} moles/liter (preferably, < 10^{-7} moles/liter). The
15   exclusion of "variable domain of an antibody" is intended to make clear that for the purposes herein a protein is not to be considered a "binding protein" merely because it is antigenic. However, an antigen may nonetheless qualify as a binding protein because it specifically
20   binds to a substance other than an antibody, e.g., an enzyme for its substrate, or a hormone for its cellular receptor. Additionally, it should be pointed out that "binding protein" may include a protein which binds specifically to the Fc of an antibody, e.g.,
25   staphylococcal protein A.

While the present invention may be used to develop novel antibodies through variegation of codons corresponding to the hypervariable region of an antibody's variable domain, its primary utility resides
30   in the development of binding proteins which are not antibodies or even variable domains of antibodies. Novel antibodies can be obtained by immunological techniques; novel enzymes, hormones, etc. cannot.

Most larger proteins fold into distinguishable
35   globules called domains. The display strategy is first perfected by modifying a genetic phage to display a

26

stable, structured domain (the "_initial potential binding_
domain", IPBD) for which an affinity molecule (which may
be an antibody) is obtainable. The success of the
modifications is readily measured by, _e.g._, determining
5   whether the modified genetic phage binds to the affinity
molecule.   For the purpose of identifying IPBDs,
definitions of "domain" which emphasize stability --
retention of the overall structure in the face of
perturbing forces such as elevated temperatures or
10   chaotropic agents -- are favored, though atomic coor-
dinates and protein sequence homology are not completely
ignored.   When a domain of a protein is primarily
responsible for the protein's ability to specifically
bind a chosen target, it is referred to herein as a
15   "binding domain" (BD).
      The IPBD is chosen with a view to its tolerance for
extensive mutagenesis.   Once it is known that the IPBD
can be displayed on a surface of a phage and subjected to
affinity selection, the gene encoding the IPBD is sub-
20   jected to a special pattern of multiple mutagenesis, here
termed "_variegation_", which after appropriate cloning and
amplification steps leads to the production of a popula-
tion of phage each of which displays a single potential
binding domain (a mutant of the IPBD), but which
25   collectively display a multitude of different though
structurally related potential binding domains (PBDs).
Each genetic phage carries the version of the _pbd_ gene
that encodes the PBD displayed on the surface of that
particular phage.   Affinity selection is then used to
30   identify the display phage bearing the PBDs with the
desired binding characteristics, and these display phage
may then be amplified. After one or more cycles of
enrichment by affinity selection and amplification, the
DNA encoding the successful binding domains (SBDs) may
35   then be recovered from selected phage.

27

If need be, the DNA from the SBD-bearing phage may
then be further "variegated", using an SBD of the last
round of variegation as the "parental potential binding
domain" (PPBD) to the next generation of PBDs, and the
5   process continued until the worker in the art is
satisfied with the result. At that point, the SBD may be
produced by any conventional means, including chemical
synthesis.

The initial potential binding domain may be: 1) a
10   domain of a naturally occurring protein, 2) a non-natur-
ally occurring domain which substantially corresponds in
sequence to a naturally occurring domain, but which
differs from it in sequence by one or more substitutions,
insertions or deletions, 3) a domain substantially
15   corresponding in sequence to a hybrid of subsequences of
two or more naturally occurring proteins, or 4) an
artificial domain designed entirely on theoretical
grounds based on knowledge of amino acid geometries and
statistical evidence of secondary structure preferences
20   of amino acids. (However, the limitations of a priori
protein design prompted the present invention.) Usually,
the domain will be a known binding domain, or at least a
homologue thereof, but it may be derived from a protein
which, while not possessing a known binding activity,
25   possesses a secondary or higher structure that lends
itself to binding activity (clefts, grooves, etc.). The
protein to which the IPBD is related need not have any
specific affinity for the target material.

In determining whether sequences should be deemed to
30   "substantially correspond", one should consider the
following issues: the degree of sequence similarity when
the sequences are aligned for best fit according to
standard algorithms, the similarity in the connectivity
patterns of any crosslinks (e.g., disulfide bonds), the
35   degree to which the proteins have similar three-dimen-
sional structures, as indicated by, e.g., X-ray diffrac-

28

tion analysis or NMR, and the degree to which the se-
quenced proteins have similar biological activity.   In
this context, it should be noted that among the serine
protease inhibitors, there are families of proteins
5   recognized to be homologous in which there are pairs of
members with as little as 30% sequence homology.

A candidate IPBD should meet the following criteria:

1) a domain exists that will remain stable under the
   conditions of its intended use (the domain may
10   comprise the entire protein that will be inserted,
   e.g. BPTI, α-conotoxin GI, or CMTI-III),

2) knowledge of the amino acid sequence is obtain-
   able, and

3) a molecule is obtainable having specific and high
15   affinity for the IPBD, AfM(IPBD).

Preferably, in order to guide the variegation strategy,
knowledge of the identity of the residues on the domain's
outer surface, and their spatial relationships, is
obtainable; however, this consideration is less important
20   if the binding domain is small, e.g., under 40 residues.

Preferably, the IPBD is no larger than necessary
because small SBDs (for example, less than 40 amino
acids) can be chemically synthesized and because it is
easier to arrange restriction sites in smaller amino-acid
25   sequences.   For PBDs smaller than about 40 residues, an
added advantage is that the entire variegated pbd gene
can be synthesized in one piece.   In that case, we need
arrange only suitable restriction sites in the osp gene.
A smaller protein minimizes the metabolic strain on the
30   phage or the host of the GP.   The IPBD is preferably
smaller than about 200 residues.   The IPBD must also be
large enough to have acceptable binding affinity and
specificity.   For an IPBD lacking covalent crosslinks,
such as disulfide bonds, the IPBD is preferably at least
35   40 residues; it may be as small as six residues if it
contains a crosslink.

29

There are many candidate IPBDs, for example, bovine
pancreatic trypsin inhibitor (BPTI, 58 residues), CMTI-
III (29 residues), crambin (46 residues), third domain of
ovomucoid (56 residues), heat-stable enterotoxin (ST-Ia
5   of E. coli) (18 residues), α-Conotoxin GI (13 residues),
μ-Conotoxin GIII (22 residues), Conus King Kong mini-
protein (27 residues), T4 lysozyme (164 residues), and
azurin (128 residues). Table 50 lists several preferred
IPBDs.

10   In some cases, a protein having some affinity for
the target may be a preferred IPBD even though some other
criteria are not optimally met. For example, the V1
domain of CD4 is a good choice as IPBD for a protein that
binds to gp120 of HIV. It is known that mutations in the
15   region 42 to 55 of V1 greatly affect gp120 binding and
that other mutations either have much less effect or
completely disrupt the structure of V1. Similarly, tumor
necrosis factor (TNF) would be a good initial choice if
one wants a TNF-like molecule having higher affinity for
20   the TNF receptor.

As even surface mutations may reduce the stability
of the PBD, the chosen IPBD should have a high melting
temperature (50°C acceptable, the higher the better; BPTI
melts at 95°C.) and be stable over a wide pH range (8.0
25   to 3.0 acceptable; 11.0 to 2.0 preferred), so that the
SBDs derived from the chosen IPBD by mutation and
selection-through-binding will retain sufficient stabil-
ity. Preferably, the substitutions in the IPBD yielding
the various PBDs do not reduce the melting point of the
30   domain below ~40°C. Mutations may arise that increase the
stability of SBDs relative to the IPBD, but the process
of the present invention does not depend upon this
occurring. Proteins containing covalent crosslinks, such
as multiple disulfides, are usually sufficiently stable.
35   A protein having at least two disulfides and having at

30

least 1 disulfide for every twenty residues may be
presumed to be sufficiently stable.

     If the target is a protein or other macromolecule a
preferred embodiment of the IPBD is a small protein such

5   as the Cucurbita maxima trypsin inhibitor III (29 resi-
dues), BPTI from Bos Taurus (58 residues), crambin from
rape seed (46 residues), or the third domain of ovomucoid
from Coturnix coturnix Japonica (Japanese quail) (56
residues), because targets from this class have clefts

10   and grooves that can accommodate small proteins in highly
specific ways. If the target is a macromolecule lacking
a compact structure, such as starch, it should be treated
as if it were a small molecule. Extended macromolecules
with defined 3D structure, such as collagen, should be

15   treated as large molecules.

     If the target is a small molecule, such as a
steroid, a preferred embodiment of the IPBD is a protein
of about 80-200 residues, such as ribonuclease from Bos
taurus (124 residues), ribonuclease from Aspergillus

20   oruzae (104 residues), hen egg white lysozyme from Gallus
gallus (129 residues), azurin from Pseudomonas aerugenosa
(128 residues), or T4 lysozyme (164 residues), because
such proteins have clefts and grooves into which the
small target molecules can fit. The Brookhaven Protein

25   Data Bank contains 3D structures for all of the proteins
listed. Genes encoding proteins as large as T4 lysozyme
can be manipulated by standard techniques for the
purposes of this invention.

     If the target is a mineral, insoluble in water, one

30   considers the nature of the molecular surface of the
mineral. Minerals that have smooth surfaces, such as
crystalline silicon, are best addressed with medium to
large proteins, such as ribonuclease, as IPBD in order to
have sufficient contact area and specificity. Minerals

35   with rough, grooved surfaces, such as zeolites, could be

bound either by small proteins, such as BPTI, or larger proteins, such as T4 lysozyme.

BPTI is an especially preferred IPBD because it meets or exceeds all the criteria: it is a small, very
5   stable protein with a well known 3D structure.

Small polypeptides have potential advantages over larger polypeptides when used as therapeutic or diagnostic agents, including (but not limited to):

    a) better penetration into tissues,
10   b) faster elimination from the circulation (important for imaging agents),

    c) lower antigenicity, and

    d) higher activity per mass.

Thus, it would be desirable to be able to employ the
15   combination of variegation and affinity selection to identify small polypeptides which bind a target of choice.

Polypeptides of this size, however, have disadvantages as binding molecules. According to Olivera et al.
20   (OLIV90a): "Peptides in this size range normally equilibrate among many conformations (in order to have a fixed conformation, proteins generally have to be much larger)." Specific binding of a peptide to a target molecule requires the peptide to take up one conformation
25   that is complementary to the binding site.

In one embodiment, the present invention overcomes these problems, while retaining the advantages of smaller polypeptides, by fostering the biosynthesis of novel mini-proteins having the desired binding characteristics.
30   Mini-Proteins are small polypeptides (usually less than about 60 residues, more preferably less than 40 residues ("micro-proteins")) which, while too small to have a stable conformation as a result of noncovalent forces alone, are covalently crosslinked (e.g., by disulfide
35   bonds) into a stable conformation and hence have biological activities more typical of larger protein

32

molecules than of unconstrained polypeptides of
comparable size.

When mini-proteins are variegated, the residues
which are covalently crosslinked in the parental molecule
5   are left unchanged, thereby stabilizing the conformation.
For example, in the variegation of a disulfide bonded
mini-protein, certain cysteines are invariant so that
under the conditions of expression and display, covalent
crosslinks (e.g., disulfide bonds between one or more
10   pairs of cysteines) form, and substantially constrain the
conformation which may be adopted by the hypervariable
linearly intermediate amino acids. In other words, a
constraining scaffolding is engineered into polypeptides
which are otherwise extensively randomized.

15   Once a mini-protein of desired binding character-
istics is characterized, it may be produced, not only by
recombinant DNA techniques, but also by nonbiological
synthetic methods.

For the purpose of the appended claims, a mini-
20   protein has between about eight and about 60 residues.
An intrachain disulfide bridge connecting amino acids 3
and 8 of a 16 residue polypeptide will be said herein to
have a span of 4. If amino acids 4 and 12 are also
disulfide bonded, then their bridge has a span of 7.
25   Together, the four cysteines divide the polypeptide into
four intercysteine segments (1-2, 5-7, 9-11, and 13-16).
(Note that there is no segment between Cys3 and Cys4.)

The connectivity pattern of a crosslinked mini-
protein is a simple description of the relative location
30   of the termini of the crosslinks. For example, for a
mini-protein with two disulfide bonds, the connectivity
pattern "1-3, 2-4" means that the first crosslinked
cysteine is disulfide bonded to the third crosslinked
cysteine (in the primary sequence), and the second to the
35   fourth.

33

The variegated disulfide-bonded mini-proteins of the
present invention fall into several classes.

Class I mini-proteins are those featuring a single
pair of cysteines capable of interacting to form a
5   disulfide bond, said bond having a span of no more than
nine residues. This disulfide bridge preferably has a
span of at least two residues; this is a function of the
geometry of the disulfide bond. When the spacing is two
or three residues, one residue is preferably glycine in
10   order to reduce the strain on the bridged residues. The
upper limit on spacing is less precise, however, in
general, the greater the spacing, the less the constraint
on conformation imposed on the linearly intermediate
amino acid residues by the disulfide bond.

15   A disulfide bridge with a span of 4 or 5 is espe-
cially preferred. If the span is increased to 6, the
constraining influence is reduced. In this case, we
prefer that at least one of the enclosed residues be an
amino acid that imposes restrictions on the main-chain
20   geometry. Proline imposes the most restriction. Valine
and isoleucine restrict the main chain to a lesser
extent. The preferred position for this constraining
non-cysteine residue is adjacent to one of the invariant
cysteines, however, it may be one of the other bridged
25   residues. If the span is seven, we prefer to include two
amino acids that limit main-chain conformation. These
amino acids could be at any of the seven positions, but
are preferably the two bridged residues that are
immediately adjacent to the cysteines. If the span is
30   eight or nine, additional constraining amino acids may be
provided.

Additional amino acids may appear on the amino side
of the first cysteine or the carboxy side of the second
cysteine. Only the immediately proximate "unspanned"
35   amino acids are likely to have a significant effect on
the conformation of the span.

34

Class II mini-proteins are those featuring a single
disulfide bond having a span of greater than nine amino
acids. The bridged amino acids form secondary structures
which help to stabilize their conformation. Preferably,
5   these   intermediate   amino   acids   form   hairpin
supersecondary   structures   such   as   those   schematized
below:

```
     ┌───────────S─S───────────┐
     -Cys-αhelix-turn-ßstrand-Cys-
```
10
```
     ┌───────────S─S───────────┐
     -Cys-αhelix-turn-αhelix-Cys-
```
```
     ┌───────────S─S───────────┐
     -Cys-ßstrand-turn-ßstrand-Cys-
```
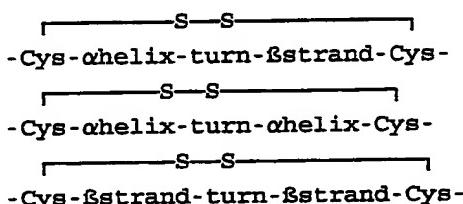
15   In designing a suitable hairpin structure, one may
copy an actual structure from a protein whose three-
dimensional conformation is known, design the structure
using   secondary   structure   tendency   data   for   the
individual amino acids, etc., or combine the two
20   approaches.   Preferably, one or more actual structures
are used as a model, and the frequency data is used to
determine which mutations can be made without disrupting
the structure.
         Preferably, no more than three amino acids lie
25   between the cysteine and the beginning or end of the α
helix or ß strand.
         More complex structures (such as a double hairpin)
are also possible.
         Class III mini-proteins are those featuring a
30   plurality of disulfide bonds.   They optionally may also
feature secondary structures such as those discussed
above with regard to Class II mini-proteins.   Since the
number of possible disulfide bond topologies increases
rapidly with the number of bonds (two bonds, three
35   topologies; three bonds, 15 topologies; four bonds, 105
topologies) the number of disulfide bonds preferably does

35

not exceed four. Two disulfide bond are preferable to three, and three to four. With two or more disulfide bonds, the disulfide bridge spans preferably do not exceed 30, and the largest intercysteine chain segment
5 preferably does not exceed 20.

Naturally occurring class III mini-proteins, such as heat-stable enterotoxin ST-Ia, frequently have pairs of cysteines that are clustered (-C-C- or -C-X-C-) in the amino-acid sequence. Clustering reduces the number of
10 realizable topologies, and may be advantageous.

Metal Finger Mini-Proteins. The mini-proteins of the present invention are not limited to those crosslinked by disulfide bonds. Another important class of mini-proteins are analogues of finger proteins.
15 Finger proteins are characterized by finger structures in which a metal ion is coordinated by two Cys and two His residues, forming a tetrahedral arrangement around it. The metal ion is most often zinc(II), but may be iron, copper, cobalt, etc. The "finger" has the consensus
20 sequence (Phe or Tyr)-(1 AA)-Cys-(2-4 AAs)-Cys-(3 AAs)-Phe-(5 AAs)-Leu-(2 AAs)-His-(3 AAs)-His-(5 AAs)(BERG88; GIBS88). The present invention encompasses mini-proteins with either one or two fingers.

Further diversity may be introduced into a display phage
25 library of potential binding domains by treating the phage with (preferably nontoxic) enzymes and/or chemical reagents that can selectively modify certain side groups of proteins, and thereby affect the binding properties of the displayed PBDs. Using affinity separation methods,
30 we enrich for the modified GPs that bind the predetermined target. Since the active binding domain is not entirely genetically specified, we must repeat the post-morphogenesis modification at each enrichment round. This approach is particularly appropriate with mini-
35 protein IPBDs because we envision chemical synthesis of these SBDs.

36

EPITOPIC PEPTIDES

The present invention also relates to the identification of epitopic peptides which bind to a target which is the epitopic binding site of an antibody,

5   lectin, enzyme, or other binding protein.   In the case of an antibody, the epitopic peptide will be at least four amino acids and more preferably at least six or eight amino acids.   Usually, it will be less than 20 amino acids, but there is no fixed upper limit.   In

10   general, however, the epitopic peptide will be a "linear" or "sequential" epitope.   Typically, in constructing a library for displaying epitopic peptides, all or most of the amino acid positions of the potential epitope will be varied.   However, it is desirable that among those amino

15   acids allowed at a particular position, that there be a relatively equal representation, as further discussed below in the context of mutagenesis of protein domains.

VARIEGATION STRATEGY -- MUTAGENESIS TO OBTAIN POTENTIAL
20   BINDING DOMAINS (OR EPITOPES) WITH DESIRED DIVERSITY

When the number of different amino acid sequences obtainable by mutation of the domain is large when compared to the number of different domains which are displayable in detectable amounts, the efficiency of the

25   forced evolution is greatly enhanced by careful choice of which residues are to be varied.   First, residues of a known protein which are likely to affect its binding activity (e.g., surface residues) and not likely to unduly degrade its stability are identified.   Then all or

30   some of the codons encoding these residues are varied simultaneously to produce a variegated population of DNA. The variegated population of DNA is used to express a variety of potential binding domains, whose ability to bind the target of interest may then be evaluated.

35   The method of the present invention is thus further distinguished from other methods in the nature of the

37

highly variegated population that is produced and from
which novel binding proteins are selected. We force the
displayed potential binding domain to sample the nearby
"sequence space" of related amino-acid sequences in an
5   efficient, organized manner. Four goals guide the
various variegation plans used herein, preferably: 1) a
very large number (e.g. 10^7) of variants is available, 2)
a very high percentage of the possible variants actually
appears in detectable amounts, 3) the frequency of
10   appearance of the desired variants is relatively uniform,
and 4) variation occurs only at a limited number of
amino-acid residues, most preferably at residues having
side groups directed toward a common region on the
surface of the potential binding domain.
15   This is to be distinguished from the simple use of
indiscriminate mutagenic agents such as radiation and
hydroxylamine to modify a gene, where there is no (or
very oblique) control over the site of mutation. Many of
the mutations will affect residues that are not a part of
20   the binding domain. Moreover, since at a reasonable
level of mutagenesis, any modified codon is likely to be
characterized by a single base change, only a limited and
biased range of possibilities will be explored. Equally
remote is the use of site-specific mutagenesis techniques
25   employing mutagenic oligonucleotides of nonrandomized
sequence, since these techniques do not lend themselves
to the production and testing of a large number of
variants. While focused random mutagenesis techniques
are known, the importance of controlling the distribution
30   of variation has been largely overlooked.
         The term "variegated DNA" (vgDNA) refers to a
mixture of DNA molecules of the same or similar length
which, when aligned, vary at some codons so as to encode
at each such codon a plurality of different amino acids,
35   but which encode only a single amino acid at other codon
positions. It is further understood that in variegated

38

DNA, the codons which are variable, and the range and
frequency of occurrence of the different amino acids
which a given variable codon encodes, are determined in
advance by the synthesizer of the DNA, even though the
5   synthetic method does not allow one to know, a priori,
the sequence of any individual DNA molecule in the
mixture. The number of designated variable codons in the
variegated DNA is preferably no more than 20 codons, and
more preferably no more than 5-10 codons. The mix of
10   amino acids encoded at each variable codon may differ
from codon to codon. A population of display phage into
which variegated DNA has been introduced is likewise said
to be "variegated".

When DNA encoding a portion of a known domain of a
15   protein is variegated, the original domain is called the
parent of the potential binding domains (PPBD), and the
multitude of mutant domains encoded as a result of the
variegation are collectively called the "potential
binding domains" (PBD), as their ability to bind to the
20   predetermined target is not then known.

We now consider the manner in which we generate a
diverse population of potential binding domains in order
to facilitate selection of a PBD-bearing phage which
binds with the requisite affinity to the target of
25   choice. The potential binding domains are first designed
at the amino acid level. Once we have identified which
residues are to be mutagenized, and which mutations to
allow at those positions, we may then design the
variegated DNA which is to encode the various PBDs so as
30   to assure that there is a reasonable probability that if
a PBD has an affinity for the target, it will be
detected. Of course, the number of independent
transformants obtained and the sensitivity of the
affinity separation technology will impose limits on the
35   extent of variegation possible within any single round of
variegation.

39

There are many ways to generate diversity in a protein. At one extreme, we vary a few residues of the protein as much as possible ("Focused Mutagenesis"), e.g., we pick a set of five to seven residues and vary
5   each through 13-20 possibilities. An alternative plan of mutagenesis ("Diffuse Mutagenesis") is to vary many more residues through a more limited set of choices (See VERS86a and PAKU86). The variegation pattern adopted may fall between these extremes.
10   There is no fixed limit on the number of codons which can be mutated simultaneously. However, it is desirable to adopt a mutagenesis strategy which results in a reasonable probability that a possible PBD sequence is in fact displayed by at least one phage. Preferably,
15   the probability that a mutein encoded by the vgDNA and composed of the least favored amino acids at each variegated position will be displayed by at least one independent transformant in the library is at least 0.50, and more preferably at least 0.90. (Muteins composed of
20   more favored amino acids would of course be more likely to occur in the same library.)
     Preferably, the variegation is such as will cause a typical transformant population to display 10^6-10^7 different amino acid sequences by means of preferably not
25   more than 10-fold more (more preferably not more than 3-fold) different DNA sequences.
     For a mini-protein that lacks α helices and ß strands, one will, in any given round of mutation, preferably variegate each of 4-6 non-cysteine codons so
30   that they each encode at least eight of the 20 possible amino acids. The variegation at each codon could be customized to that position. Preferably, cysteine is not one of the potential substitutions, though it is not excluded.
35   When the mini-protein is a metal finger protein, in a typical variegation strategy, the two Cys and two His

40

residues, and optionally also the aforementioned Phe/Tyr,
Phe and Leu residues, are held invariant and a plurality
(usually 5-10) of the other residues are varied.

5   When the mini-protein is of the type featuring one
or more α helices and ß strands, the set of potential
amino acid modifications at any given position is picked
to favor those which are less likely to disrupt the
secondary structure at that position. Since the number
of possibilities at each variable amino acid is more
10   limited, the total number of variable amino acids may be
greater without altering the sampling efficiency of the
selection process.

For the last-mentioned class of mini-proteins, as
well as domains other than mini-proteins, preferably not
15   more than 20 and more preferably 5-10 codons will be
variegated. However, if diffuse mutagenesis is employed,
the number of codons which are variegated can be higher.

The decision as to which residues to modify is eased
by knowledge of which residues lie on the surface of the
20   domain and which are buried in the interior.

We choose residues in the PPBD to vary through
consideration of several factors, including: a) the 3D
structure of the PPBD, b) sequences homologous to PPBD,
and c) modeling of the PPBD and mutants of the PPBD.
25   When the number of residues that could strongly influence
binding without preventing the normal folding of the PPBD
is greater than the number that should be varied
simultaneously, the user should pick a subset of those
residues to vary at one time. The user picks trial
30   levels of variegation and calculate the abundances of
various sequences. The list of varied residues and the
level of variegation at each varied residue are adjusted
until the composite variegation is commensurate with the
sensitivity of the affinity separation and the number of
35   independent transformants that can be made.

41

Having picked which residues to vary, we now decide the range of amino acids to allow at each variable residue. The total level of variegation is the product of the number of variants at each varied residue. Each varied residue can have a different scheme of variegation, producing 2 to 20 different possibilities. The set of amino acids which are potentially encoded by a given variegated codon are called its "substitution set".

The computer that controls a DNA synthesizer, such as the Milligen 7500, can be programmed to synthesize any base of an oligo-nt with any distribution of nts by taking some nt substrates (e.g. nt phosphoramidites) from each of two or more reservoirs. Alternatively, nt substrates can be mixed in any ratios and placed in one of the extra reservoir for so called "dirty bottle" synthesis. Each codon could be programmed differently. The "mix" of bases at each nucleotide position of the codon determines the relative frequency of occurrence of the different amino acids encoded by that codon.

Simply variegated codons are those in which those nucleotide positions which are degenerate are obtained from a mixture of two or more bases mixed in equimolar proportions. These mixtures are described in this specification by means of the standardized "ambiguous nucleotide" code. In this code, for example, in the degenerate codon "SNT", "S" denotes an equimolar mixture of bases G and C, "N", an equimolar mixture of all four bases, and "T", the single invariant base thymidine.

Complexly variegated codons are those in which at least one of the three positions is filled by a base from an other than equimolar mixture of two of more bases.

Either simply or complexly variegated codons may be used to achieve the desired substitution set.

If we have no information indicating that a particular amino acid or class of amino acid is appropriate,

42

we strive to substitute all amino acids with equal
probability because representation of one mini-protein
above the detectable level is wasteful. Equal amounts of
all four nts at each position in a codon (NNN) yields the
5   amino acid distribution in which each amino acid is
present in proportion to the number of codons that code
for it. This distribution has the disadvantage of giving
two basic residues for every acidic residue.   In
addition, six times as much R, S, and L as W or M occur.
10   If five codons are synthesized with this distribution,
each of the 243 sequences encoding some combination of L,
R, and S are 7776-times more abundant than each of the 32
sequences encoding some combination of W and M. To have
five Ws present at detectable levels, we must have each
15   of the (L,R,S) sequences present in 7776-fold excess.
        It is generally accepted that the sequence of amino
acids in a protein or polypeptide determine the three-
dimensional structure of the molecule, including the
possibility of no definite structure. Among polypeptides
20   of definite length and sequence, some have a defined
tertiary structure and most do not.
        Particular amino acid residues can influence the
tertiary structure of a defined polypeptide in several
ways, including by:
25   a) affecting the flexibility of the polypeptide main
             chain,
          b) adding hydrophobic groups,
          c) adding charged groups,
          d) allowing hydrogen bonds, and
30   e) forming cross-links, such as disulfides, chelation
             to metal ions, or bonding to prosthetic groups.
Flexibility:
        GLY is the smallest amino acid, having two hydrogens
attached to the C_α. Because GLY has no C_β, it confers the
35   most flexibility on the main chain. Thus GLY occurs very

43

frequently in reverse turns, particularly in conjunction
with PRO, ASP, ASN, SER, and THR.

   The amino acids ALA, SER, CYS, ASP, ASN, LEU, MET,
PHE, TYR, TRP, ARG, HIS, GLU, GLN, and LYS have
5   unbranched ß carbons. Of these, the side groups of SER,
ASP, and ASN frequently make hydrogen bonds to the main
chain and so can take on main-chain conformations that
are energetically unfavorable for the others. VAL, ILE,
and THR have branched ß carbons which makes the extended
10   main-chain conformation more favorable. Thus VAL and ILE
are most often seen in ß sheets. Because the side group
of THR can easily form hydrogen bonds to the main chain,
it has less tendency to exist in a ß sheet.

   The   main   chain   of   proline   is   particularly
15   constrained by the cyclic side group. The ϕ angle is
always close to -60°. Most prolines are found near the
surface of the protein.

Charge:

   LYS and ARG carry a single positive charge at any pH
20   below 10.4 or 12.0, respectively. Nevertheless, the
methylene groups, four and three respectively, of these
amino acids are capable of hydrophobic interactions. The
guanidinium group of ARG is capable of donating five
hydrogens simultaneously, while the amino group of LYS
25   can donate only three. Furthermore, the geometries of
these groups is quite different, so that these groups are
often not interchangeable.

   ASP and GLU carry a single negative charge at any pH
above ≈4.5 and 4.6, respectively. Because ASP has but
30   one methylene group, few hydrophobic interactions are
possible. The geometry of ASP lends itself to forming
hydrogen bonds to main-chain nitrogens which is
consistent with ASP being found very often in reverse
turns and at the beginning of helices. GLU is more often
35   found in α helices and particularly in the amino-terminal
portion of these helices because the negative charge of

44

the side group has a stabilizing interaction with the
helix dipole (NICH88, SALI88).

HIS has an ionization pK in the physiological range,
viz. 6.2.   This pK can be altered by the proximity of
5   charged groups or of hydrogen donators or acceptors. HIS
is capable of forming bonds to metal ions such as zinc,
copper, and iron.

Hydrogen bonds:

Aside from the charged amino acids, SER, THR, ASN,
10 GLN, TYR, and TRP can participate in hydrogen bonds.

Cross links:

The most important form of cross link is the disul-
fide bond formed between two thiols, especially the
thiols of CYS residues. In a suitably oxidizing environ-
15 ment, these bonds form spontaneously. These bonds can
greatly stabilize a particular conformation of a protein
or mini-protein. When a mixture of oxidized and reduced
thiol reagents are present, exchange reactions take place
that allow the most stable conformation to predominate.
20 Concerning disulfides in proteins and peptides, see also
KATZ90, MATS89, PERR84, PERR86, SAUE86, WELL86, JANA89,
HORV89, KISH85, and SCHN86.

Other cross links that form without need of specific
enzymes include:

25   1) $(CYS)_4$:Fe   Rubredoxin (in CREI84, P.376)

2) $(CYS)_4$:Zn   Aspartate   Transcarbamylase   (in
CREI84,   P.376)   and   Zn-fingers
(HARD90)

3) (H I S)$_2$ (M E T) (C Y S) : C u
30   Azurin (in CREI84, P.376) and Basic "Blue" Cu
Cucumber protein (GUSS88)

4) $(HIS)_4$:Cu   CuZn superoxide dismutase

5) $(CYS)_4$:(Fe_4S_4) Ferredoxin (in CREI84, P.376)

6) $(CYS)_2(HIS)_2$:Zn Zinc-fingers (GIBS88)

35   7) $(CYS)_3(HIS)$:Zn Zinc-fingers (GAUS87, GIBS88)

45

Cross links having (HIS)₂(MET)(CYS):Cu has the potential
advantage that HIS and MET can not form other cross links
without Cu.

Simply Variegated Codons

5   The following simply variegated codons are useful
because they encode a relatively balanced set of amino
acids:

1) SNT which encodes the set [L,P,H,R,V,A,D,G]: a) one
   acidic (D) and one basic (R), b) both aliphatic
10   (L,V) and aromatic hydrophobics (H), c) large
   (L,R,H) and small (G,A) side groups, d) ridged (P)
   and flexible (G) amino acids, e) each amino acid
   encoded once.

2) RNG which encodes the set [M,T,K,R,V,A,E,G]: a) one
15   acidic and two basic (not optimal, but acceptable),
   b) hydrophilics and hydrophobics, c) each amino acid
   encoded once.

3) RMG which encodes the set [T,K,A,E]: a) one acidic,
   one basic, one neutral hydrophilic, b) three favor
20   α helices, c) each amino acid encoded once.

4) VNT which encodes the set [L,P,H,R,I,T,N,S,V,A,D,G]:
   a) one acidic, one basic, b) all classes: charged,
   neutral hydrophilic, hydrophobic, ridged and flex-
   ible, etc., c) each amino acid encoded once.

25   5) RRS which encodes the set [N,S,K,R,D,E,G²]: a) two
   acidics, two basics, b) two neutral hydrophilics, c)
   only glycine encoded twice.

6) N N T   w h i c h   e n c o d e s   t h e   s e t
   [F,S,Y,C,L,P,H,R,I,T,N,V,A,D,G]:   a) sixteen DNA
30   sequences provide fifteen different amino acids;
   only serine is repeated, all others are present in
   equal amounts (This allows very efficient sampling
   of the library.), b) there are equal numbers of
   acidic and basic amino acids (D and R, once each),
35   c) all major classes of amino acids are present:

46

acidic, basic, aliphatic hydrophobic, aromatic
hydrophobic, and neutral hydrophilic.

7) N N G ,   w h i c h   e n c o d e s   t h e   s e t
   [L², R², S, W, P, Q, M, T, K, V, A, E, G,   stop]:   a)   fair
5 preponderance of residues that favor formation of α-
   helices [L, M, A, Q, K, E;   and,   to   a   lesser   extent,
   S, R, T]; b) encodes 13 different amino acids. (VHG
   encodes a subset of the set encoded by NNG which
   encodes   9   amino   acids   in   nine   different   DNA
10 sequences, with equal acids and bases, and 5/9 being
   α helix-favoring.)

   For the initial variegation, NNT is preferred, in
most cases. However, when the codon is encoding an amino
acid to be incorporated into an α helix, NNG is
15 preferred.

   Below, we analyze several simple variegations as to
the efficiency with which the libraries can be sampled.

   Libraries of random hexapeptides encoded by (NNK)[6]
have been reported (SCOT90, CWIR90). Table 130 shows the
20 expected behavior of such libraries. NNK produces single
codons for PHE, TYR, CYS, TRP, HIS, GLN, ILE, MET, ASN,
LYS, ASP, and GLU (α set); two codons for each of VAL,
ALA, PRO, THR, and GLY (Φ set); and three codons for each
of LEU, ARG, and SER (Ω set). We have separated the
25 64,000,000 possible sequences into 28 classes, shown in
Table 130A, based on the number of amino acids from each
of these sets. The largest class is ΦΩααα with ≈14.6%
of the possible sequences. Aside from any selection, all
the sequences in one class have the same probability of
30 being produced. Table 130B shows the probability that a
given DNA sequence taken from the (NNK)[6] library will
encode a hexapeptide belonging to one of the defined
classes; note that only ≈6.3% of DNA sequences belong to
the ΦΩαααα class.

47

Table 130C shows the expected numbers of sequences
in each class for libraries containing various numbers of
independent transformants (<u>viz.</u> 10^6, $3 \cdot 10^6$, 10^7, $3 \cdot 10^7$, 10^8,
$3 \cdot 10^8$, 10^9, and $3 \cdot 10^9$). At 10^6 independent transformants

5   (ITs), we expect to see 56% of the ΩΩΩΩΩΩ class, but only
0.1% of the αααααα class. The vast majority of sequences
seen come from classes for which less than 10% of the
class is sampled. Suppose a peptide from, for example,
class ΦΦΩΩαα is isolated by fractionating the library for

10   binding to a target. Consider how much we know about
peptides that are related to the isolated sequence.
Because only 4% of the ΦΦΩΩαα class was sampled, we can
not conclude that the amino acids from the Ω set are in
fact the best from the Ω set. We might have LEU at

15   position 2, but ARG or SER could be better. Even if we
isolate a peptide of the ΩΩΩΩΩΩ class, there is a notice-
able chance that better members of the class were not
present in the library.

With a library of 10^7 ITs, we see that several

20   classes have been completely sampled, but that the αααααα
class is only 1.1% sampled. At $7.6 \cdot 10^7$ ITs, we expect
display of 50% of all amino-acid sequences, but the
classes containing three or more amino acids of the α set
are still poorly sampled. To achieve complete sampling

25   of the $(NNK)^6$ library requires about $3 \cdot 10^9$ ITs, 10-fold
larger than the largest $(NNK)^6$ library so far reported.

Table 131 shows expectations for a library encoded
by $(NNT)^4(NNG)^2$. The expectations of abundance are
independent of the order of the codons or of interspersed

30   unvaried codons. This library encodes 0.133 times as
many amino-acid sequences, but there are only 0.0165
times as many DNA sequences. Thus $5.0 \cdot 10^7$ ITs (<u>i.e.</u> 60-
fold fewer than required for $(NNK)^6$) gives almost complete
sampling of the library. The results would be slightly

35   better for $(NNT)^6$ and slightly, but not much, worse for

48

(NNG)[6]. The controlling factor is the ratio of DNA sequences to amino-acid sequences.

Table 132 shows the ratio of #DNA sequences/#AA sequences for codons NNK, NNT, and NNG. For NNK and NNG,
5   we have assumed that the PBD is displayed as part of an essential gene, such as gene III in Ff phage, as is indicated by the phrase "assuming stops vanish". It is not in any way required that such an essential gene be used. If a non-essential gene is used, the analysis
10   would be slightly different; sampling of NNK and NNG would be slightly less efficient. Note that (NNT)[6] gives 3.6-fold more amino-acid sequences than (NNK)[5] but requires 1.7-fold fewer DNA sequences. Note also that (NNT)[7] gives twice as many amino-acid sequences as (NNK)[6],
15   but 3.3-fold fewer DNA sequences.

Thus, while it is possible to use a simple mixture (NNS, NNK or NNN) to obtain at a particular position all twenty amino acids, these simple mixtures lead to a highly biased set of encoded amino acids. This problem
20   can be overcome by use of complexly variegated codons.

We first will present the mixture calculated (see WO90/02809) to minimize the ratio of most favored amino acid to least favored amino acid when the nt distribution is subject to two constraints: equal abundances of
25   acidic and basic amino acids and the least possible number of stop codons. We have simplified the search for an optimal nt distribution by limiting the third base to T or G (C or G is equivalent). However, it should be noted that the present invention embraces use of
30   complexly variegated codons in which the third base is not limited to T or G (or to C or G).

The optimum distribution (the "fxS" codon) is shown in Table 10A and yields DNA molecules encoding each type amino acid with the abundances shown. Note that this
35   chemistry encodes all twenty amino acids, with acidic and

49

basic amino acids being equiprobable, and the most
favored amino acid (serine) is encoded only 2.454 times
as often as the least favored amino acid (tryptophan).
The "fxS" vg codon improves sampling most for peptides
5   containing several of the amino acids [F,Y,C,W,H-
,Q,I,M,N,K,D,E] for which NNK or NNS provide only one
codon. Its sampling advantages are most pronounced when
the library is relatively small.

The results of searhing only for the complexly
10   variegated codon which minimizes the ratio of most
favored to least favored amino acid, without additional
constraints, is shown in Table 10B. The changes are
small, indicating that insisting on equality of acids and
bases and minimizing stop codons costs us little. Also
15   note that, without restraining the optimization, the
prevalence of acidic and basic amino acids comes out
fairly close. On the other hand, relaxing the
restriction leaves a distribution in which the least
favored amino acid is only .412 times as prevalent as
20   SER.

The advantages of an NNT codon are discussed else-
where in the present application. Unoptimized NNT
provides 15 amino acids encoded by only 16 DNA sequences.
It is possible to improve on NNT with the complexly
25   variegated codon shown in Table 10C. This gives five
amino acids (SER, LEU, HIS, VAL, ASP) in very nearly
equal amounts. A further eight amino acids (PHE, TYR,
ILE, ASN, PRO, ALA, ARG, GLY) are present at 78% the
abundance of SER. THR and CYS remain at half the abun-
30   dance of SER. When variegating DNA for disulfide-bonded
mini-proteins, it is often desirable to reduce the
prevalence of CYS. This distribution allows 13 amino
acids to be seen at high level and gives no stops; the
optimized fxS distribution allows only 11 amino acids at
35   high prevalence.

50

The NNG codon can also be optimized. When equimolar
T,C,A,G are used in NNG, one obtains double doses of LEU
and ARG. Table 10D shows an approximately optimized NNG
codon. There are, under this variegation, four equally
5   most favored amino acids: LEU, ARG, ALA, and GLU. Note
that there is one acidic and one basic amino acid in this
set. There are two equally least favored amino acids:
TRP and MET. The ratio of lfaa/mfaa is 0.5258. If this
codon is repeated six times, peptides composed entirely
10   of TRP and MET are 2% as common as peptides composed
entirely of the most favored amino acids. We refer to
this as "the prevalence of (TRP/MET)^6 in optimized NNG^6
vgDNA".

When synthesizing vgDNA by the "dirty bottle"
15   method, it is sometimes desirable to use only a limited
number of mixes. One very useful mixture is called the
"optimized NNS mixture" in which we average the first two
positions of the fxS mixture: $T_1 = 0.24$, $C_1 = 0.17$, $A_1 =$
0.33, $G_1 = 0.26$, the second position is identical to the
20   first, $C_3 = G_3 = 0.5$. This distribution provides the
amino acids ARG, SER, LEU, GLY, VAL, THR, ASN, and LYS at
greater than 5% plus ALA, ASP, GLU, ILE, MET, and TYR at
greater than 4%.

An additional complexly variegated codon is of
25   interest. This codon is identical to the optimized NNT
codon at the first two positions and has T:G::90:10 at
the third position. This codon provides thirteen amino
acids (ALA, ILE, ARG, SER, ASP, LEU, VAL, PHE, ASN, GLY,
PRO, TYR, and HIS) at more than 5.5%. THR at 4.3% and
30   CYS at 3.9% are more common than the LFAAs of NNK
(3.125%). The remaining five amino acids are present at
less than 1%. This codon has the feature that all amino
acids are present; sequences having more than two of the
low-abundance amino acids are rare. When we isolate an
35   SBD using this codon, we can be reasonably sure that the

first 13 amino acids were tested at each position. A
similar codon, based on optimized NNG, could be used.

    Several   of   the   preferred   simple   or   complex
variegated codons encode a set of amino acids which
5   includes cysteine. This means that some of the encoded
binding domains will feature one or more cysteines in
addition to the invariant disulfide-bonded cysteines.
For example, at each NNT-encoded position, there is a one
in sixteen chance of obtaining cysteine. If six codons
10   are   so   varied,   the   fraction   of   domains   containing
additional cysteines is 0.33. Odd numbers of cysteines
can lead to complications, see Perry and Wetzel (PERR84).
On the other hand, many disulfide-containing proteins
contain cysteines that do not form disulfides, e.g.
15   trypsin. The possibility of unpaired cysteines can be
dealt with in several ways:

    First, the variegated phage population can be passed
over an immobilized reagent that strongly binds free
thiols, such as SulfoLink (catalogue number 44895 H from
20   Pierce   Chemical   Company,   Rockford,   Illinois,   61105).
Another product from Pierce is TNB-Thiol Agarose (Cata-
logue   Code   20409   H).   BioRad   sells   Affi-Gel   401
(catalogue 153-4599) for this purpose.

    Second, one can use a variegation that excludes
25   cysteines, such as:

    NHT that gives [F,S,Y,L,P,H,I,T,N,V,A,D],

    VNS that gives

        [L^2,P^2,H,Q,R^3,I,M,T^2,N,K,S,V^2,A^2,E,D,G^2],

    NNG that gives [L^3,S,W,P,Q,R^2,M,T,K,R,V,A,E,G,stop],

30   SNT that gives [L,P,H,R,V,A,D,G],

    RNG that gives [M,T,K,R,V,A,E,G],

    RMG that gives [T,K,A,E],

    VNT that gives [L,P,H,R,I,T,N,S,V,A,D,G], or

    RRS that gives [N,S,K,R,D,E,G^2].

35   However, each of these schemes has one or more of the
disadvantages, relative to NNT: a) fewer amino acids are

52

allowed, b) amino acids are not evenly provided, c)
acidic and basic amino acids are not equally likely), or
d) stop codons occur. Nonetheless, NNG, NHT, and VNT are
almost as useful as NNT. NNG encodes 13 different amino
5   acids and one stop signal. Only two amino acids appear
twice in the 16-fold mix.

        Thirdly, one can enrich the population for binding
to the preselected target, and evaluate selected
sequences post hoc for extra cysteines. Those that
10   contain more cysteines than the cysteines provided for
conformational constraint may be perfectly usable. It is
possible that a disulfide linkage other than the designed
one will occur. This does not mean that the binding
domain defined by the isolated DNA sequence is in any way
15   unsuitable. The suitability of the isolated domains is
best determined by chemical and biochemical evaluation of
chemically synthesized peptides.

        Lastly, one can block free thiols with reagents,
such as Ellman's reagent, iodoacetate, or methyl iodide,
20   that specifically bind free thiols and that do not react
with disulfides, and then leave the modified phage in the
population. It is to be understood that the blocking
agent may alter the binding properties of the mini-
protein; thus, one might use a variety of blocking
25   reagent in expectation that different binding domains
will be found. The variegated population of thiol-
blocked display phage are fractionated for binding. If
the DNA sequence of the isolated binding mini-protein
contains an odd number of cysteines, then synthetic means
30   are used to prepare mini-proteins having each possible
linkage and in which the odd thiol is appropriately
blocked. Nishiuchi (NISH82, NISH86, and works cited
therein) disclose methods of synthesizing peptides that
contain a plurality f cysteines so that each thiol is
35   protected with a different type of blocking group. These
groups can be selectively removed so that the disulfide

53

pairing can be controlled. We envision using such a
scheme with the alteration that one thiol either remains
blocked, or is unblocked and then reblocked with a
different reagent.

5   Use of NNT or NNG variegated codons leads to very
efficient sampling of variegated libraries because the
ratio of (different amino-acid sequences)/(different DNA
sequences) is much closer to unity than it is for NNK or
even the optimized vg codon (fxS). Nevertheless, a few
10   amino acids are omitted in each case. Both NNT and NNG
allow members of all important classes of amino acids:
hydrophobic, hydrophilic, acidic, basic, neutral
hydrophilic, small, and large. After selecting a binding
domain, a subsequent variegation and selection may be
15   desirable to achieve a higher affinity or specificity.
During this second variegation, amino acid possibilities
overlooked by the preceding variegation may be
investigated.
        In the second round of variegation, a preferred
20   strategy is to vary each position through a new set of
residues which includes the amino acid(s) which were
found at that position in the successful binding domains,
and which include as many as possible of the residues
which were excluded in the first round of variegation.
25   Thus, later rounds of variegation test both amino
acid positions not previously mutated, and amino acid
substitutions at a previously mutated position which were
not within the previous substitution set.
        If the first round of variegation is entirely
30   unsuccessful, a different pattern of variegation should
be used. For example, if more than one interaction set
can be defined within a domain, the residues varied in
the next round of variegation should be from a different
set than that probed in the initial variegation. If
35   repeated failures are encountered, one may switch to a
different IPBD.

54

AFFINITY SELECTION OF TARGET-BINDING MUTANTS

Affinity separation is used initially in the present invention to verify that the display system is working, i.e., that a chimeric outer surface protein has been
5   expressed and transported to the surface of the phage and is oriented so that the inserted binding domain is accessible to target material. When used for this purpose, the binding domain is a known binding domain for a particular target and that target is the affinity
10   molecule used in the affinity separation process. For example, a display system may be validated by using inserting DNA encoding BPTI into a gene encoding an outer surface protein of the phage of interest, and testing for binding to anhydrotrypsin, which is normally bound by
15   BPTI.

If the phage bind to the target, then we have confirmation that the corresponding binding domain is indeed displayed by the phage. Phage which display the binding domain (and thereby bind the target) are
20   separated from those which do not.

Once the display system is validated, it is possible to use a variegated population of phage which display a variety of different potential binding domains, and use affinity separation technology to determine how well they
25   bind to one or more targets. This target need not be one bound by a known binding domain which is parental to the displayed binding domains, i.e., one may select for binding to a new target.

For example, one may variegate a BPTI binding domain
30   and test for binding, not to trypsin, but to another serine protease, such as human neutrophil elastase or cathepsin G, or even to a wholly unrelated target, such as horse heart myoglobin.

The term "affinity separation means" includes, but
35   is not limited to: a) affinity column chromatography, b) batch elution from an affinity matrix material, c) batch

55

elution from an affinity material attached to a plate, d)
fluorescence activated cell sorting, and e) electrophor-
esis in the presence of target material. "Affinity
material" is used to mean a material with affinity for
5   the material to be purified, called the "analyte". In
most cases, the association of the affinity material and
the analyte is reversible so that the analyte can be
freed from the affinity material once the impurities are
washed away.

10   If affinity chromatography is to be used, then:
      1) the molecules of the target material must be of
         sufficient size and chemical reactivity to be
         applied to a solid support suitable for affinity
         separation,
15   2) after application to a matrix, the target material
         preferably does not react with water,
      3) after application to a matrix, the target material
         preferably does not bind or degrade proteins in a
         non-specific way, and
20   4) the molecules of the target material must be suffi-
         ciently large that attaching the material to a
         matrix allows enough unaltered surface area (gener-
         ally at least 500 $Å^2$, excluding the atom that is
         connected to the linker) for protein binding.

25   Affinity chromatography is the preferred separation
means, but FACS, electrophoresis, or other means may also
be used.

          The present invention makes use of affinity separa-
tion of phage to enrich a population for those phage
30   carrying genes that code for proteins with desirable
binding properties.

          The present invention may be used to select for
binding domains which bind to one or more target mater-
ials, and/or fail to bind to one or more target
35   materials. Specificity, of course, is the ability of a
binding molecule to bind strongly to a limited set of

56

target materials, while binding more weakly or not at all
to another set of target materials from which the first
set must be distinguished.

Almost any molecule that is suitable for affinity
5   separation may be used as a target. Possible targets
include, but are not limited to peptides, soluble and
insoluble proteins, nucleic acids, lipids, carbohydrates,
other organic molecules (monomeric or polymeric),
inorganic compounds, and organometallic compounds.
10   Serine proteases are an especially interesting class of
potential target materials.

For chromatography, FACS, or electrophoresis there
may be a need to covalently link the target material to
a second chemical entity. For chromatography the second
15   entity is a matrix, for FACS the second entity is a
fluorescent dye, and for electrophoresis the second
entity is a strongly charged molecule. In many cases, no
coupling is required because the target material already
has the desired property of: a) immobility, b) fluores-
20   cence, or c) charge. In other cases, chemical or
physical coupling is required.

It is not necessary that the actual target material
be used in preparing the immobilized or labeled analogue
that is to be used in affinity separation; rather,
25   suitable reactive analogues of the target material may be
more convenient. Target materials that do not have
reactive functional groups may be immobilized by first
creating a reactive functional group through the use of
some powerful reagent, such as a halogen. In some cases,
30   the reactive groups of the actual target material may
occupy a part on the target molecule that is to be left
undisturbed. In that case, additional functional groups
may be introduced by synthetic chemistry.

Two very general methods of immobilization are
35   widely used. The first is to biotinylate the compound of
interest and then bind the biotinylated derivative to

57

immobilized avidin. The second method is to generate
antibodies to the target material, immobilize the anti-
bodies by any of numerous methods, and then bind the
target material to the immobilized antibodies. Use of
5 antibodies is more appropriate for larger target materi-
als; small targets (those comprising, for example, ten or
fewer non-hydrogen atoms) may be so completely engulfed
by an antibody that very little of the target is exposed
in the target-antibody complex.

10   Non-covalent immobilization of hydrophobic molecules
without resort to antibodies may also be used. A com-
pound, such as 2,3,3-trimethyldecane is blended with a
matrix precursor, such as sodium alginate, and the
mixture is extruded into a hardening solution. The
15 resulting beads will have 2,3,3-trimethyldecane dispersed
throughout and exposed on the surface.

Other immobilization methods depend on the presence
of particular chemical functionalities. A polypeptide
will present -NH$_2$ (N-terminal; Lysines), -COOH (C-ter-
20 minal; Aspartic Acids; Glutamic Acids), -OH (Serines;
Threonines; Tyrosines), and -SH (Cysteines). A polysac-
charide has free -OH groups, as does DNA, which has a
sugar backbone.

The following table is a nonexhaustive review of
25 reactive functional groups and potential immobilization
reagents:

Group	Reagent
R-NH$_2$	Derivatives of 2,4,6-trinitro benzene sulfonates (TNBS), (CREI84, p.11)
R-NH$_2$	Carboxylic acid anhydrides, _e.g._ derivatives of succinic anhydride, maleic anhydride, citraconic anhydride (CREI84, p.11)
R-NH$_2$	Aldehydes that form reducible Schiff bases (CREI84, p.12) guanido cyclohexanedione derivatives (CREI84, p.14)

58

R-CO$_2$H	Diazo cmpds (CREI84, p.10)
R-CO$_2$-	Epoxides (CREI84, p.10)
R-OH	Carboxylic acid anhydrides
Aryl-OH	Carboxylic acid anhydrides
Indole ring	Benzyl halide and sulfenyl halides (CREI84, p.19)
R-SH	N-alkylmaleimides (CREI84, p.21)
R-SH	ethyleneimine derivatives (CREI84, p.21)
R-SH	Aryl mercury compounds, (CREI84, P.21)
R-SH	Disulfide reagents, (CREI84, p.23)
Thiol ethers	Alkyl iodides, (CREI84, p.20) Ketones Make Schiff's base and reduce with NaBH$_4$. (CREI84, p.12-13)
Aldehydes	Oxidize to COOH, vide supra.
R-SO$_3$H	Convert to R-SO$_2$Cl and react with immobilized alcohol or amine.
R-PO$_3$H	Convert to R-PO$_2$Cl and react with immobilized alcohol or amine.
CC double bonds	Add HBr and then make amine or thiol.

The extensive literature on affinity chromatography and related techniques will provide further examples.

Matrices suitable for use as support materials include polystyrene, glass, agarose and other chromatographic supports, and may be fabricated into beads, sheets, columns, wells, and other forms as desired. Suppliers of support material for affinity chromatography include: Applied Protein Technologies Cambridge, MA; Bio-Rad Laboratories, Rockville Center, NY; Pierce Chemical Company, Rockford, IL. Target materials are attached to the matrix in accord with the directions of the manufacturer of each matrix preparation with consideration of good presentation of the target.

59

Early in the selection process, relatively high
concentrations of target materials may be applied to the
matrix to facilitate binding; target concentrations may
subsequently be reduced to select for higher affinity
5   SBDs.

The population of display phage is applied to an
affinity matrix under conditions compatible with the
intended use of the binding protein and the population is
fractionated by passage of a gradient of some solute over
10 the column.   The process enriches for PBDs having
affinity for the target and for which the affinity for
the target is least affected by the eluants used.   The
enriched fractions are those containing viable display
phage that elute from the column at greater concentration
15 of the eluant.

The eluants preferably are capable of weakening
noncovalent interactions between the displayed PBDs and
the immobilized target material.   Preferably, the eluants
do not kill the phage; the genetic message corresponding
20 to   successful   mini-proteins   is   most   conveniently
amplified by reproducing the phage rather than by _in_
vitro procedures such as PCR.   The list of potential
eluants includes salts (including Na+, NH$_4$+, Rb+, SO$_4$--,
H$_2$PO$_4$-, citrate, K+, Li+, Cs+, HSO$_4$-, CO$_3$--, Ca++, Sr++,
25 Cl-, PO$_4$---, HCO$_3$-, Mg++, Ba++, Br-, HPO$_4$-- and acetate),
acid,   heat,   compounds known to bind the target, and
soluble target material (or analogues thereof).

Neutral solutes, such as ethanol, acetone, ether, or
urea, are frequently used in protein purification and are
30 known   to   weaken   non-covalent   interactions   between
proteins and other molecules. Many of these species are,
however,   very harmful   to bacteria and bacteriophage.
Urea is known not to harm M13 up to 8 M.   Salt is a
preferred solute for gradient formation in most cases.
35 Decreasing pH is also a highly preferred eluant.   In some
cases, the preferred matrix is not stable to low pH so

60

that salt and urea are the most preferred reagents.
Other solutes that generally weaken non-covalent interac-
tion between proteins and the target material of interest
may also be used.

5   The uneluted display phage contain DNA encoding
binding domains which have a sufficiently high affinity
for the target material to resist the elution conditions.
The DNA encoding such successful binding domains may be
recovered in a variety of ways. Preferably, the bound
10   display phage are simply eluted by means of a change in
the elution conditions. Alternatively, one may culture
the phage in situ, or extract the target-containing
matrix with phenol (or other suitable solvent) and
amplify the DNA by PCR or by recombinant DNA techniques.
15   Or, if a site for a specific protease has been engineered
into the display vector, the specific protease is used to
cleave the binding domain from the GP.

         Variation in the support material (polystyrene,
glass, agarose, cellulose, etc.) in analysis of clones
20   carrying SBDs is used to distinguish phage that bind to
the support material rather than the target.

         The harvested phage are now enriched for the
binding-to-target phenotype by use of affinity separation
involving the target material immobilized on an affinity
25   matrix. Phage that fail to bind to the target material
are washed away. It may be desirable to include a
bacteriocidal agent, such as azide, in the buffer to
prevent bacterial growth. The buffers used in
chromatography include: a) any ions or other solutes
30   needed to stabilize the target, and b) any ions or other
solutes needed to stabilize the PBDs derived from the
IPBD.

         Recovery of phage that display binding to an
affinity column is typically achieved by collecting
35   fractions eluted from the column with a gradient of a
chaotropic agent as described above, or of the target

61

material in soluble form; fractions eluting later in the
gradient are enriched for high-affinity phage.   The
eluted phage are then amplified in suitable host cells.

If some high-affinity phage cannot be eluted from
5   the target in viable form, one may:
   1) flood the matrix with a nutritive medium and grow
      the desired phage in situ,
   2) remove parts of the matrix and use them to inoculate
      growth medium,
10   3) chemically or enzymatically degrade the linkage
      holding the target to the matrix so that GPs still
      bound to target are eluted, or
   4) degrade the phage and recover DNA with phenol or
      other suitable solvent; the recovered DNA is used to
15   transform cells that regenerate GPs.
It is possible to utilize combinations of these methods.
It should be remembered that what we want to recover from
the affinity matrix is not the phage per se, but the
information in them as to the sequence of the successful
20   epitope or binding domain.

As described in WO90/02809, one may modify the
affinity separation of the method described to select a
molecule that binds to material A but not to material B,
or that binds to both A and B at competing or
25   noncompeting sites, or that do not bind to selected
targets.
SUBSEQUENT PRODUCTION

Using the method of the present invention, we can
obtain a replicable phage that displays a novel protein
30   domain having high affinity and specificity for a target
material of interest.   Such a phage carries both amino-
acid embodiments of the binding protein domain and a DNA
embodiment of the gene encoding the novel binding domain.
The presence of the DNA facilitates expression of a
35   protein comprising the novel binding protein domain

62

within a high-level expression system, which need not be
the same system used during the developmental process.

        We can proceed to production of the novel binding
protein in several ways, including: a) altering of the
5   gene encoding the binding domain so that the binding
domain is expressed as a soluble protein, not attached to
a phage (either by deleting codons 5' of those encoding
the binding domain or by inserting stop codons 3' of
those encoding the binding domain), b) moving the DNA
10   encoding the binding domain into a known expression
system, and c) utilizing the phage as a purification
system.   (If the domain is small enough, it may be
feasible to prepare it by conventional peptide synthesis
methods.)

15   As previously mentioned, an advantage inhering from
the use of a mini-protein as an IPBD is that it is likely
that the derived SBD will also behave like a mini-protein
and will be obtainable by means of chemical synthesis.
(The term "chemical synthesis", as used herein, includes
20   the use of enzymatic agents in a cell-free environment.)

        Peptides may be chemically synthesized either in
solution or on supports.   Various combinations of
stepwise synthesis and fragment condensation may be
employed.

25   During synthesis, the amino acid side chains are
protected to prevent branching.   Several different
protective groups are useful for the protection of the
thiol groups of cysteines:

        1) 4-methoxybenzyl (MBzl; Mob)(NISH82; ZAFA88), remov-
30   able with HF;

        2) acetamidomethyl (Acm)(NISH82; NISH86; BECK89c),
           removable with iodine; mercury ions (e.g., mercuric
           acetate); silver nitrate; and

        3) S-para-methoxybenzyl.

63

Other thiol protective groups may be found in standard reference works such as Greene, PROTECTIVE GROUPS IN ORGANIC SYNTHESIS (1981).

5   Once the polypeptide chain has been synthesized, disulfide bonds must be formed.   Possible oxidizing agents include air (NISH86), ferricyanide (NISH82), iodine (NISH82), and performic acid.   Temperature, pH, solvent, and chaotropic chemicals may affect the course of the oxidation.

10   A large number of mini-proteins with a plurality of disulfide bonds have been chemically synthesized in biologically active form.

The successful binding domains of the present invention may, alone or as part of a larger protein, be
15   used for any purpose for which binding proteins are suited, including isolation or detection of target materials.   In furtherance of this purpose, the novel binding proteins may be coupled directly or indirectly, covalently or noncovalently, to a label, carrier or
20   support.

When used as a pharmaceutical, the novel binding proteins may be contained with suitable carriers or adjuvanants.

* * * * *

25   All references cited anywhere in this specification are incorporated by reference to the extent which they may be pertinent.

64

All cells used in the following examples are E. coli
cells.

EXAMPLE I

DISPLAY OF BPTI AS A FUSION TO M13 GENE VIII PROTEIN:

5   Example I involves display of BPTI on M13 as a
fusion to the mature gene VIII coat protein. Each DNA
construction was confirmed by restriction digestion and
DNA sequencing.

1.   Construction   of   the   viii-signal-
10   sequence::bpti::mature-viii-coat-protein Display Vector.

The operative cloning vectors are M13 and phagemids
derived from M13. The initial construction was in the
f1-based phagemid pGEM-3Zf(-)^(TM) (Promega Corp., Madison,
WI.).

15   We constructed a gene encoding, in order,: i) a
modified lacUV5 promoter, ii) a Shine-Dalgarno sequence,
iii) M13 gene VIII signal sequence, iv) mature BPTI, v)
mature-M13-gene-VIII coat protein, vi) multiple stop
codons, and vii) a transcription terminator. This gene
20   is illustrated in Table 102. The operator of lacUV5 is
the symmetrical lacO to allow tighter repression in the
absence of IPTG. The longest segment that is identical
to wild-type gene VIII is minimized so that genetic
recombination with the co-existing gene VIII is unlikely.
25   i) OCV based upon pGEM-3Zf.

pGEM-3Zf^(TM) (Promega Corp., Madison, WI.) is a vector
containing the amp gene, bacterial origin of replication,
bacteriophage f1 origin of replication, a lacZ operon
containing a multiple cloning site sequence, and the T7
30   and SP6 polymerase binding sequences.

BamHI and SalI sites were introduced at the
boundaries of the lacZ operon (to facilitate removal of
the lacZ operon and its replacement with the synthetic
gene); this vector is named pGEM-MB3/4.

35   ii) OCV based upon M13mp18.

65

M13mp18 (YANI85) is a vector (New England Biolabs,
Beverly, MA.) consisting of the whole of the phage genome
plus a lacZ operon containing a multiple cloning site
(MESS77). BamHI and SalI sites were introduced into
5   M13mp18 at the 5' and 3' ends of the lacZ operon; this
vector is named M13-MB1/2.

B) Synthetic Gene.

A synthetic gene (VIII-signal-sequence::mature-
bpti::mature-VIII-coat-protein) was constructed from 16
10   synthetic oligonucleotides, synthesized by Genetic
Designs Inc. of Houston, Texas, via a method similar to
those in KIMH89 and ASHM89. Table 102 contains an
annotated version of this sequence. The oligonucleotides
were phosphorylated, with the exception of the 5' most
15   molecules, using standard methods, annealed and ligated
in stages. The overhangs were filled in with T4 DNA
polymerase and the DNA was cloned into the HincII site of
pGEM-3Zf(-); the initial construct is pGEM-MB1. Double-
stranded DNA of pGEM-MB1 was cut with PstI, filled in
20   with T4 DNA polymerase and ligated to a SalI linker (New
England BioLabs) so that the synthetic gene is bounded by
BamHI and SalI sites (Table 102). The synthetic gene was
obtained on a BamHI-SalI cassette and cloned into pGEM-
MB3/4 and M13-MB1/2 using the introduced BamHI and SalI
25   sites, to generate pGEM-MB16 and M13-MB15, respectively.
The synthetic insert was sequenced. The original
Ribosome Binding Site (RBS) was in error (AGAGG instead
of the designed AGGAGG) and we detected no expressed
protein in vivo and in vitro.

30   C) Alterations to the synthetic gene.

i) Ribosome binding site (RBS).

In pGEM-MB16, a SacI-NheI DNA fragment (containing
the RBS) was replaced with an oligonucleotide encoding a
new RBS very similar to the RBS of E. coli phoA that is
35   known to function.

66

Original putative RBS (5'-to-3')

GAGCTCagaggCTTACTATGAAGAAATCTCTGGTTCTTAAGGCTAGC
|SacI| | Nhe I |

5   New RBS (5'-to-3')

GAGCTCTggaggaAATAAAATGAAGAAATCTCTGGTTCTTAAGGCTAGC
|SacI| | Nhe I |

10   The putative RBSs are lower case and the initiating
     methionine codon is underscored and bold. The resulting
     construct is pGEM-MB20. In vitro expression of the gene
     carried by pGEM-MB20 produced a novel protein species of
     the expected size, about 14.5 kd.

15   ii) tac promoter.
          To obtain higher expression of the fusion protein,
     the lacUV5 promoter was changed to a tac promoter. In
     pGEM-MB16, a BamHI-HpaII fragment (containing the lacUV5
     promoter) was replaced with an oligonucleotide containing
20   the -35 sequence of the trp promoter (Cf RUSS82)
     converting the lacUV5 promoter to tac. The vector is
     named pGEM-MB22.

     MB16   5'- GATCC tctagagtcggc TTTACA ctttatgcttc(cg-
25   gctcg..-3'
               3'-   G agatctcagccg aaatgt gaaatacgaag
     gc(cgagc..-5'
               | | | -35| | |
               BamHI   HpaII

30

     MB22   5'- GATCC actccccatccccctg TTGACA attaatcat -3'
             3'-   G tgaggggtaggggggac AACTGT taattagtagc-5'
               | | | -35| |
35   BamHI   (HpaII)

          Promoter and RBS variants of the fusion protein gene
     were constructed as follows:

40   _____Promoter___RBS_____Encoded Protein._____

     pGEM-MB16   lac   old   VIIIs.p.-BPTI-matureVIII
     pGEM-MB20   lac   new   ''
     pGEM-MB22   tac   old   ''
45   pGEM-MB26   tac   new   ''

67

The synthetic genes from pGEM-MB20 and pGEM-MB26 were recloned into the altered phage vector M13-MB1/2 to generate the phage M13-MB27 and M13-MB28 respectively.

iii. Signal Peptide Sequence.

5   In vitro expression of the synthetic gene regulated by tac and the "new" RBS produced a novel protein of the expected size for the unprocessed protein (~16 kd). In vivo expression also produced novel protein of full size; no processed protein could be seen on phage or in cell
10   extracts by silver staining or by Western analysis with anti-BPTI antibody.

Thus we analyzed the signal sequence of the fusion. Table 106 shows a number of typical signal sequences. Charged residues are generally thought to be of great
15   importance and are shown **bold** and underscored. Each signal sequence contains a long stretch of uncharged residues that are mostly hydrophobic; these are shown in lower case. At the right, in parentheses, is the length of the stretch of uncharged residues. We note that the
20   fusions of gene VIII signal to BPTI and gene III signal to BPTI have rather short uncharged segments. These short uncharged segments may reduce or prevent processing of the fusion peptides. We know that the gene III signal sequence is capable of directing: a) insertion of the
25   peptide   comprising   (mature-BPTI)::(mature-gene-III-protein) into the lipid bilayer, and b) translocation of BPTI and most of the mature gene III protein across the lipid bilayer (vide infra). That the gene III remains anchored   in   the   lipid   bilayer   until   the   phage   is
30   assembled is directed by the uncharged anchor region near the carboxy terminus of the mature gene III protein (see Table 116) and not by the secretion signal sequence. The phoA signal sequence can direct secretion of mature BPTI into the periplasm of E. coli (MARK86). Furthermore,
35   there is controversy over the mechanism by which mature

68

authentic gene VIII protein comes to be in the lipid
. bilayer prior to phage assembly.

Thus we replaced the DNA coding for the gene-VIII-
putative-signal-sequence by each of DNA coding for: 1)
5   the phoA signal sequence, 2) the bla signal sequence, and
3) the M13 gene III signal. Each of these replacements
produces a tripartite gene encoding a fusion protein that
comprises, in order: (a) a signal peptide that directs
secretion into the periplasm of parts (b) and (c),
10   derived from a first gene; (b) an initial potential
binding domain (BPTI in this case), derived from a second
gene (in this case, the second gene is an animal gene);
and (c) a structural packaging signal (the mature gene
VIII coat protein), derived from a third gene.

15   The process by which the IPBD::packaging-signal fusion
arrives on the phage surface is illustrated in Figure 1.
In Figure 1a, we see that authentic gene VIII protein
appears (by whatever process) in the lipid bilayer so
that both the amino and carboxy termini are in the cyto-
20   plasm. Signal peptidase-I cleaves the gene VIII protein
liberating the signal peptide (that is absorbed by the
cell) and mature gene VIII coat protein that spans the
lipid bilayer. Many copies of mature gene VIII coat
protein accumulate in the lipid bilayer awaiting phage
25   assembly (Figure 1c). Some signal sequences are able to
direct the translocation of quite large proteins across
the lipid bilayer. If additional codons are inserted
after the codons that encode the cleavage site of the
signal peptidase-I of such a potent signal sequence, the
30   encoded amino acids will be translocated across the lipid
bilayer as shown in Figure 1b. After cleavage by signal
peptidase-I, the amino acids encoded by the added codons
will be in the periplasm but anchored to the lipid
bilayer by the mature gene VIII coat protein, Figure 1d.
35   The circular single-stranded phage DNA is extruded
through a part of the lipid bilayer containing a high

69

concentration of mature gene VIII coat protein; the
carboxy terminus of each coat protein molecule packs near
the DNA while the amino terminus packs on the outside.
Because the fusion protein is identical to mature gene
5   VIII coat protein within the trans-bilayer domain, the
fusion protein will co-assemble with authentic mature
gene VIII coat protein as shown in Figure 1e.

In each case, the mature VIII coat protein moiety is
intended to co-assemble with authentic mature VIII coat
10   protein to produce phage particle having BPTI domains
displayed on the surface. The source and character of
the secretion signal sequence is not important because
the signal sequence is cut away and degraded.   The
structural packaging signal, however, is quite important
15   because it must co-assemble with the authentic coat
protein to make a working virus sheath.

a) Bacterial Alkaline Phosphatase (phoA) Signal Peptide.

Construct pGEM-MB26 contains a SacI-AccIII fragment
containing the new RBS and sequences encoding the
20   initiating methionine and the signal peptide of M13 gene
VIII pro-protein.   This fragment was replaced with a
duplex (annealed from four oligo-nts) containing the RBS
and DNA coding for the initiating methionine and signal
peptide of PhoA (INOU82); phage is pGEM-MB42. M13MB48 is
25   a derivative of GemMB42.   A BamHI-SalI DNA fragment from
GenMB42, containing the gene construct, was ligated into
a similarly cleaved vector M13MB1/2 giving rise to
M13MB48.

PhoA RBS and signal peptide sequence

70

5'-GAGCTCCATGGGAGAAAATAAA.ATG.AAA.CAA.AGC.ACG.-
   |SacI| met lys gln ser thr

     .ATC.GCA.CTC.TTA.CCG.TTA.CTG.TTT.ACC.CCT.GTG.ACA.-
5   ile ala leu leu pro leu leu phe thr pro val thr

          .AAA.GCC.CGT.CCG.GAT.-3'
          lys ala arg pro asp......
                   |AccIII|
10

b) β-lactamase signal peptide.

   To allow transfer of the β-lactamase (amp) promoter and
DNA coding for the signal peptide into the (mature-
BPTI)::(mature-VIII-coat-protein) gene, we first
15 introduced an AccIII site the amp gene adjacent to the
codons for the β-lactamase signal peptide cleavage site
(C_{2504}→T and A_{2501}→G); vector is pGEM-MB40. We then ligated
a BamHI linker into the AatII site at nt 2260, 5' to the
promoter; vector is pGEM-MB45. The BamHI-AccIII fragment
20 now contains the amp promoter, amp RBS, initiating
methionine and β-lactamase signal peptide. This fragment
was used to replace the corresponding fragment from pGEM-
MB26 to generate construct pGEM-MB46.

25   amp gene promoter and signal peptide sequences

5'-GGATCCGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTT-

     TATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACC-
30
     CTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGT-

     ATG.AGT.ATT.CAA.CAT.TTC.CGT.GTC.GCC.CTT.ATT.CCC.TTT.TTT.-
35   met ser ile gln his phe arg val ala leu ile pro phe
     phe

     GCG.GCA.TTT.TGC.CTT.CCT.GTT.TTT.GCT.CAT.CCG.-3'
     ala ala phe cys leu pro val phe ala his pro....
40

c) M13-gene-III-signal::bpti::mature-VIII-coat-protein

   We may also construct M13-MB51 which would carry a Km^R
gene and a M13-gene-III-signal-peptide gene fragment
fused to the previously described BPTI::mature-VIII-coat-
45 protein gene fragment. Because M13-MB51 contains no gene

71

III, the phage can not form plaques, but can render cells Km^R. Infectious phage particles can be obtained via helper phage. The gene III signal sequence is capable of directing (BPTI)::(mature-gene-III-protein) to the surface of phage (vide infra).

Summary of signal peptide fusion protein variants.

	Promoter	RBS	Signal sequence	Fusion protein
pGEM-MB26	tac	new	VIII	BPTI/VIII-coat
pGEM-MB42	tac	new	phoA	BPTI/VIII-coat
pGEM-MB46	amp	amp	amp	BPTI/VIII-coat
pGEM-MB51 (hypoth.)	III	III	III	BPTI/VIII-coat
M13 MB48	tac	new	phoA	BPTI/VIII-coat

2. Analysis of the Protein Products Encoded by the Synthetic (signal-peptide::mature-bpti::viii-coat-protein) Genes

i) In vitro analysis

A coupled transcription/translation prokaryotic system (Amersham Corp., Arlington Heights, IL) was utilized for the in vitro analysis of the protein products encoded by the BPTI/VIII synthetic gene and the derivatives.

Table 107 lists the protein products encoded by the listed vectors which are visualized by standard fluorography following in vitro synthesis in the presence of ^{35}S-methionine and separation of the products using SDS polyacrylamide gel electrophoresis. In each sample, a pre-β-lactamase product ($\approx$31 kd) can be seen. This is derived from the amp gene which is the common selection gene for the vectors. In addition, a (pre-BPTI/VIII) product encoded by the synthetic gene and variants can be seen as indicated. The migration of these species ($\approx$14.5 kd) is consistent with the expected size of the encoded proteins.

ii) In vivo analysis.

72

The vectors detailed in sections (B) and (C) were freshly transfected into the E. coli strain XL1-blue ™ (Stratagene, La Jolla, CA) and in strain SEF'. E. coli strain SE6004 (LISS85) carries the prlA4 mutation and is

5   more permissive in secretion than strains that carry the wild-type prlA. SE6004 is F⁻ and is deleted for lacI; thus the cells can not be infected by M13 and lacUV5 and tac promoters can not be regulated with IPTG. Strain SEF' is derived from strain SE6004 (LISS85) by crossing with XL1-

10   Blue™; the F' in XL1-Blue™ carries Tc^R and lacI^q. SE6-004 is streptomycin^R, Tc^S while XL1-Blue™ is strepto-mycin^S, Tc^R so that both parental strains can be killed with the combination of Tc and streptomycin. SEF' retains the secretion-permissive phenotype of the

15   parental strain, SE6004(prlA4).

The fresh transfectants were grown in NZYCM medium (SAMB89) for 1 hour. IPTG was added over the range 1.0 μM to 0.5 mM (to derepress lacUV5 and tac) and grown for an additional 1.5 hours.

20   Aliquots of cells expressing the synthetic-insert encoded proteins together with controls (no vector, mock vector, and no IPTG) were lysed in SDS gel loading buffer and electrophoresed in 20% polyacrylamide gels containing SDS and urea. Duplicate gels were either silver stained

25   (Daiichi, Tokyo, Japan) or electrotransferred to a nylon matrix (Immobilon from Millipore, Bedford, MA) for western analysis using rabbit anti-BPTI polyclonal antibodies.

Table 108 lists the interesting proteins seen by silver

30   staining and western analysis of identical gels. We can see clearly by western analysis that IPTG-inducible protein species containing BPTI epitopes exist in the test strains which are absent from the control strains. In XL1-Blue™, the migration of this species is predomin-

35   antly that of the unprocessed form of the pro-protein although a small proportion of the protein appears to

73

migrate at a size consistent with that of a fully processed form.   In SEF', the processed form predominates, there being only a faint band corresponding to the unprocessed species.

5   Thus, in strain SEF', we have produced a tripartite fusion protein that is specifically cleaved after the secretion signal sequence. We believe that the mature protein comprises BPTI followed by the gene VIII coat protein and that the coat protein moiety spans the

10   membrane.   One or more copies, perhaps hundreds of copies, of this protein will co-assemble into M13 derived phage or M13-like phagemids. This construction will allow us to a) mutagenize the BPTI domain, b) display each of the variants on the coat of one or more phage

15   (one type per phage), and c) recover those phage that display variants having novel binding properties with respect to target materials of our choice.

        Rasched and Oberer (RASC86) report that phage produced in cells that express two alleles of gene VIII, that have

20   differences within the first 11 residues of the mature coat protein, contain some of each protein.   Thus, because we have achieved in vivo processing of the phoA(signal)::bpti::matureVIII fusion gene, it is highly likely that co-expression of this gene with wild-type

25   VIII will lead to production of phage bearing BPTI domains on their surface. Mutagenesis of the bpti domain of these genes will provide a population of phage, each phage carrying a gene that codes for the variant of BPTI displayed on the phage surface.

30   **VIII Display Phage: Production, Preparation and Analysis.**
        i. Phage Production.
        The OCV can be grown in XL1-Blue$^{(TM)}$ in the absence of IPTG. Typically, a plaque plug is taken from a plate and grown in 2 ml of medium, containing freshly diluted

35   cells, for 6 to 8 hours. Following centrifugation of this culture, the supernatant is titered. This is kept

74

as a phage stock for further infection, phage production, and display of the gene product of interest.

A 100-fold dilution of a fresh overnight culture of SEF' cells in 500 ml of NZCYM medium is allowed to grow to a
5   cell density of 0.4 (Ab 600nm) in a shaker at 37°C. To this culture is added a sufficient amount of the phage stock to give a MOI of 10 together with IPTG to give a final concentration of 0.5 mM. The culture is allowed to grow for a further 2 hrs.

10   **ii. Phage Preparation and Purification.**

The phage-producing bacterial culture is centrifuged to separate the phage in the supernatant from the bacterial pellet. To the supernatant is added one quarter, by volume, of phage precipitation solution (20% PEG, 3.75 M
15   ammonium acetate) and PMSF to a final concentration of 1mM. It is left on ice for 2 hours after which the precipitated phage is retrieved by centrifugation. The phage pellet is redissolved in TrisEDTA containing 0.1% Sarkosyl and left at 4°C for 1 hour after which any
20   bacteria and bacterial debris is removed by centrifuga- tion. The phage in the supernatant is reprecipitated with PEG overnight at 4°C. The phage pellet is resuspended in LB medium and repreciptated another two times to remove the detergent. The phage is stored in LB
25   medium at 4°C, titered, and used for analysis and binding studies.

A more stringent phage purification scheme involves centrifugation in a CsCl gradient (3.86 g of CsCl dissolved in NET buffer (0.1 M NaCl, 1mM EDTA, 0.1M Tris
30   pH 7.7) to make 10 ml). 10^{12} to 10^{13} phage in TE Sarkosyl buffer are mixed with 5 ml of CsCl NET buffer and centrifuged overnight at 34K rpm in, for example, a Sorvall OTD-65B Ultracentrifuge. Aliquots of 400 μl are carefully removed. 5 μl aliqouts are removed from the
35   fractions and analysed by agarose gel electrophoresis after heating at 65°C for 15 minutes together with the

75

gel loading buffer containing 0.1% SDS.   Fractions
containing phage are pooled, the phage reprecipitated and
finally redissolved in LB medium to a concentration of
10^{12} to 10^{13} phage per ml.

5   **iii. Phage Analysis.**

The display phage are analyzed using standard methods
of polyacrylamide gel electrophoresis and either silver
staining of the gel or electrotransfer to a nylon matrix
followed by analysis with anti-BPTI antiserum (Western
10   analysis).   Display of heterologous proteins is
quantitated by comparison to serial dilutions of the
starting protein, for example BPTI, together with the
display phage samples in the electrophoresis and Western
analyses.   An alternative method involves running a 2-
15   fold serial dilution of a phage in which both the major
coat protein and the fusion protein are silver stained.
Comparison of the ratios of the two protein species
allows one to estimate the number of fusion proteins per
phage since the number of VIII gene encoded proteins per
20   phage (~3000) is known.

Incorporation of fusion protein into bacteriophage.

In vivo expression of the processed BPTI:VIII fusion
protein, encoded by vectors GemMB42 and M13MB48,
indicated that the processed fusion product is probably
25   located within the cell membrane.   Thus, it could be
incorporated into the phage and that the BPTI moiety
would be displayed at the phage surface.

SEF' cells were infected with either M13MB48 or M13mp18,
as control.   The resulting phage were electrophoresed (~
30   10^{11} phage per lane) in a 20% polyacrylamide gel
containing urea followed by electrotransfer to a nylon
matrix and western analysis using anti-BPTI rabbit serum.
A single species of protein was observed in phage derived
from infection with the M13MB48 stock phage which was not
35   observed in the control infection.   This protein migrated

76

at an apparent size of ~12 kd, consistent with that of
the fully processed fusion protein.

Western analysis of SEF' bacterial lysate with or
without phage infection demonstrated another species of
5   protein of about 20kd. This species was also present, to
a lesser degree, in phage preparations which were simply
PEG precipitated without further purification (for
example, using nonionic detergent or by CsCl gradient
centrifugation).   A   comparison   of   M13MB48   phage
10   preparations made in the presence or absence of detergent
aldemonstrated that sarkosyl treatment and CsCl gradient
purification did remove the bacterial contaminant while
having no effect on the presence of the BPTI:VIII fusion
protein. This indicates that the fusion protein has been
15   incorporated and is a constituent of the phage body.

The time course of phage production and BPTI:VIII
incorporation was followed post-infection and after IPTG
induction.   Phage   production   and   fusion   protein
incorporation appeared to be maximal after two hours.
20   This   time   course   was   utilized   in   further   phage
productions and analyses.

Polyacrylamide electrophoresis of the phage prepara-
tions, followed by silver staining, demonstrated that the
preparations   were   essentially   free   of   contaminating
25   protein species and that an extra protein band was
present in M13MB48 derived phage which was not present in
the control phage.   The size of the new protein was
consistent with that seen by western analysis. A similar
analysis of a serially diluted BPTI:VIII incorporated
30   phage demonstrated that the ratio of fusion protein to
major coat protein was typically about 1:150. Since the
phage   contains   about   3000   copies   of   the   gene   VIII
product, the phage population contains, on average, 10's
of copies of fusion protein per phage.
35   **Altering the initiating methionine of the natural gene
VIII.**

77

The OCV M13MB48 contains the synthetic gene encoding the
BPTI:VIII fusion protein in the intergenic region of the
modified M13mp18 phage vector.   The remainder of the
vector consists of the M13 genome.   To increase the phage
5   incorporation of the fusion protein, we decided to
diminish the production of the natural gene VIII product
by altering the initiating methionine codon of this gene
to CTG.   In such cases, methionine is incorporated, but
the rate of initiation is reduced.   The change was
10   achieved by site-specific oligonucleotide mutagenesis.

```
                              M    K    K    S    -rest of VIII
                    ACT.TCC.TC.ATG.AAA.AAG.TCT.
        rest of XI -  T    S    S stop
```
15
Site-specific mutagenesis.

```
                           (L)   K    K    S    -rest of
VIII
20                 ACT.TCC.AG.CTG.AAA.AAG.TCT.
         rest of XI - T    S    S  stop
```

Analyses of the phage derived from this modified vector
indicated that there was a significant increase in the
25   ratio of   fusion   protein   to   major   coat   protein.
Quantitative estimates indicated that within a phage
population as much as 100 copies of the BPTI:VIII fusion
were incorporated per phage.
Display of BPTI:VIII fusion protein by bacteriophage.
30   The BPTI:VIII fusion protein had been shown to be
incorporated into the body of the phage. This phage was
analyzed further to demonstrate that the BPTI moiety was
accessible to specific antibodies and hence displayed at
the phage surface.
35   We added purified polyclonal rabbit anti-BPTI IgG to a
known titer of phage. Following incubation, protein A-
agarose beads are added to bind IgG and left to incubate
overnight.   The IgG-protein A beads and any bound phage
are removed by centrifugation followed by a retitering of
40   the supernatant to determine loss of phage.   The phage

78

bound to the beads can also be acid eluted and titered.
The assay includes controls, such as a WT phage stock
(M13mp18) and IgG purified from normal rabbit pre-immune
serum.

5   Table 140 shows that while the titer of the WT phage is
unaltered by anti-BPTI IgG, BPTI-IIIMK (positive control,
vide infra), demonstrated a significant drop in titer
with or without the extra addition of protein A beads.
(Note, since the BPTI moiety is part of gIIIp that binds

10   phage to bacterial pili, this is expected.) Two batches
of M13MB48 phage (containing the BPTI:VIII fusion
protein) demonstrated a significant reduction in titer,
as judged by pfus, when anti-BPTI antibodies and protein
A beads were added. The initial drop in titer with the

15   antibody alone, differs somewhat between the two batches
of phage. Retrieval of the immunoprecipitated phage,
while not quantitative, was significant when compared to
the WT phage control.

       Further controls are shown in Table 141 and Table 142.

20   The data demonstrated that the loss in titer observed for
the BPTI:VIII containing phage is a result of the display
of BPTI epitopes by these phage and the specific
interaction with anti-BPTI antibodies. No significant
interaction with either protein A agarose beads or IgG

25   purified from normal rabbit serum could be demonstrated.
The larger drop in titer for M13MB48 batch five reflects
the higher level incorporation of the fusion protein in
this preparation.

       Functionality of the BPTI moiety in the BPTI-VIII display

30   **phage.**

       The previous two sections demonstrated that the
BPTI:VIII fusion protein has been incorporated into the
phage body and that the BPTI moiety is displayed at the
phage surface. To demonstrate that the displayed

35   molecule is functional, binding experiments were
performed in a manner almost identical to that described

79

in the previous section except that proteases were used
in place of antibodies. The display phage and controls
are allowed to interact with immobilized proteases or
immobilized inactivated proteases. Binding is assessed
5   by the loss in titer of the display phage or by
determining the phage bound to the beads.

Table 143 shows the results of an experiment in which
BPTI.VIII display phage, M13MB48, were allowed to bind to
anhydrotrypsin-agarose beads.   There was a significant
10   drop in titer compared to WT phage (no displayed BPTI).
A pool of phage (5AA Pool), each contain a variegated 5
amino acid extension at the BPTI:major coat protein
interface, demonstrated a similar decline in titer.   In
control (table 143), very little non-specific binding of
15   the display phage was observed with agarose beads
carrying an unrelated protein (streptavidin).

Actual binding of the display phage is demonstrated by
the data shown for two experiments in Table 144. The
negative control is M13mp18 and the positive control is
20   BPTI-IIIMK, a phage in which the BPTI moiety, attached to
the gene III protein, has been shown to be displayed and
functional.   M13MB48 and M13MB56 both bind to
anhydrotrypsin beads in a manner comparable to that of
the positive control, being 40 to 60 times better than
25   the negative control (non-display phage).   Hence,
functionality of the BPTI moiety, in the major coat
fusion protein, was established.

Furthermore, Table 145 compares binding to active and
inactivated trypsin by phage. The control phage, M13mp18
30   and BPTI-III MK, demonstrated binding similar to that
detailed elsewhere in the present application. Note that
the relative binding is enhanced with trypsin due to the
apparent marked reduction in the non-specific binding of
the WT phage to the active protease.   M13.3X7 and
35   M13.3X11, which each contain 'EGGGS' linker extensions at
the domain interface, bound to anhydrotrypsin and trypsin

80

in a manner similar to BPTI-IIIMK phage. The binding, relative to non-display phage, was ~100-fold higher in the anhydrotrypsin binding assay and at least 1000-fold higher in the trypsin binding assay. The binding of
5 another 'EGGGS' linker variant (M13.3Xd) was similar to that of M13.3X7.

To demonstrate the specificity of binding the assays were repeated with human neutrophil elastase (HNE) beads and compared to that seen with trypsin beads Table 146.
10 BPTI has a very high affinity for trypsin and a low affinity for HNE, hence the BPTI display phage should reflect these affinities when used in binding assays with these beads. The negative and positive controls for trypsin binding were as already described above while an
15 additional positive control for the HNE beads, BPTI(K15L,MGNG)-III MA was included. The results, shown in Table 146, confirmed this prediction. M13MB48, M13.3X7 and M13.3X11 phage demonstrated good binding to trypsin, relative to WT phage and the HNE control
20 (BPTI(K15L,MGNG)-III MA), being comparable to BPTI-IIIMK phage. Conversely poor binding occurred when HNE beads were used, with the exception of the HNE positive control phage.

Taken together the accumulated data demonstrated that
25 when BPTI is part of a fusion protein with the major coat protein of M13 phage, the molecule is both displayed at the surface of the phage and a significant proportion of it is functional in a specific protease binding manner.

--- *** ---

30   EXAMPLE II
        CONSTRUCTION OF BPTI/GENE-III DISPLAY VECTOR

DNA manipulations were conducted according to standard procedures as described in Maniatis et al. (MANI82) or Sambrook et al. (SAMB89). First the lacZ gene of M13-
35 MB1/2 was removed. M13-MB1/2 RF was cut with BamHI and SalI and the large fragment was isolated. The recovered

81

6819 bp fragment was filled in with Klenow enzyme, ligated to a HindIII 8mer linker, and used to transfect XL1-Blue^(TM) (Stratagene, La Jolla, CA) cells which were subsequently plated for plaque formation.   RF DNA was

5   prepared from chosen plaques and a clone, M13-MB1/2--delta, containing regenerated BamHI and SalI sites and a new HindIII site, all 500 bp upstream of the BglII site (6935), was picked.

     A unique NarI site was introduced into codons 17 and 18

10   of gene III (changing the amino acids from H-S to G-A, Cf. Table 110) in M13-MB1/2-delta:

```
      13   14   15   16   17   18   19   20   21
      P    F    Y    S    H    S    A    E    T
      5'-ct ttc tat tct cac tcc gct gaa ac-3' wild-type
15    3'-ga aag ata aga ccg cgg cga ctt tg-5'
mutagenizing oligo
      5'-ct ttc tat tct ggc gcc gct gaa ac-3' mutant
        P    F    Y    S    G    A    A    E    T
```

The presence of a unique NarI site at nucleotide 1630 was

20   confirmed by restriction enzyme analysis; the new vector is M13-MB1/2-delta-NarI.   Phage MK was made by cloning the 1.3 Kb BamHI Km^R fragment from plasmid pUC4K (Pharmacia, Piscataway, NJ) into M13-MB1/2-delta-NarI. Phage MK grows as well as wild-type M13, indicating that

25   the changes at the cleavage site of gene III protein are not detectably deleterious to the phage.

INSERTION OF SYNTHETIC BPTI GENE

     The BPTI-III expression vector was constructed by standard means.   The synthetic bpti-VIII fusion contains

30   a NarI site that comprises the last two codons of the BPTI-encoding region.   A second NarI site was introduced upstream of the BPTI-encoding region by ligating the adaptor shown to AccIII-cut M13-MB26:

```
      5'-TATTCTGGCGCCCGT      -3'
35    3'-ATAAGACCGCGGGCAGGCC-5'
              |NarI|  |AccIII
```

The ligation sample was then restricted with NarI and a 180 bp DNA fragment encoding BPTI was isolat d.   RF DNA

82

of phage MK was digested with NarI, dephosphorylated, and
ligated to the 180 bp fragment. Ligation samples were
used to transfect XL1-Blue™ which were plated for Km^R
plaques. DNA, isolated from phage derived from plaques

5   was test for hybridization to a ^{32}P-phosphorylated double
stranded DNA probe corresponding to the BPTI gene. Large
scale RF preparations were made for clones exhibiting a
strong hybridization signal. Restriction enzyme
digestion analysis confirmed the insertion of a single

10 copy of the synthetic BPTI gene into gene III of MK to
generate phage MK-BPTI. Subsequent DNA sequencing con-
firmed that the sequence of the bpti-III fusion gene is
correct and that the correct reading frame is maintained.
Table 116 shows the entire coding region, the translation

15 into protein sequence, and the functional parts of the
polypeptide chain.

EXPRESSION OF THE BPTI-III FUSION GENE IN VITRO

MK-BPTI RF DNA was added to a coupled prokaryotic
transcription-translation extract (Amersham). Newly

20 synthesized radiolabelled proteins were produced and
subsequently separated by electrophoresis on a 15% SDS-
polyacrylamide gel. The MK-BPTI DNA directs the
synthesis of an unprocessed gene III fusion protein which
is 7 Kd larger than the WT gene III, consistent with

25 insertion of 58 amino acids of BPTI into gene III
protein. We immunoprecipitated radiolabelled proteins
from the cell-free prokaryotic extract. Neither rabbit
anti(M13-gene-VIII-protein) IgG nor normal rabbit IgG
were able to immunoprecipitate the gene III protein

30 encoded by either MK or MK-BPTI. However, rabbit
anti-BPTI IgG is able to precipitate the gene III protein
encoded by MK-BPTI but not by MK. This confirms that the
increase in size of the III protein encoded by MK-BPTI is
attributable to the insertion of the BPTI protein.

35 WESTERN ANALYSIS

83

Phage were recovered from cultures by PEG precipitation. To remove residual cells, recovered phage were resuspended in a high salt buffer and centrifuged, as per instructions for the MUTA-GENE(R) M13 in vitro Mutagenesis

5   Kit (Catalogue Number 170-3571, Bio-Rad, Richmond, CA). Aliquots of phage (containing up to 40 μg of protein) were electrophoresed on a 12.5% SDS-urea-polyacrylamide gel and proteins were electro-transferred to a sheet of Immobilon.   Western blots were developed using rabbit

10 anti-BPTI serum, which had previously been incubated with an E. coli extract, followed by goat ant-rabbit antibody conjugated to alkaline phosphatase.   An immunoreactive protein of 67 Kd is detected in preparations of the MK-BPTI but not the MK phage.   The size of the immunore-

15 active protein is consistent with the predicted size of a processed BPTI-III fusion protein (6.4 Kd plus 60 Kd). These data indicate that BPTI-specific epitopes are presented on the surface of the MK-BPTI phage but not the MK phage.

20 **NEUTRALIZATION OF PHAGE TITER WITH AGAROSE-IMMOBILIZED ANHYDRO-TRYPSIN**

Anhydro-trypsin is a derivative of trypsin having the active site serine converted to dehydroalanine. Anhydro-trypsin retains the specific binding of trypsin but not

25 the protease activity.   Unlike polyclonalantibodies, anhydro-trypsin is not expected to bind unfolded BPTI or incomplete fragments.

Phage MK-BPTI and MK were diluted to a concentration $1.4 \cdot 10^{12}$ particles per ml. in TBS buffer (PARM88) contain-

30 ing 1.0 mg/ml BSA. 30 μl of diluted phage were added to 2, 5, or 10 μl of a 50% slurry of agarose-immobilized anhydro-trypsin (Pierce Chemical Co., Rockford, IL) in TBS/BSA buffer.   Following incubation at 25°C, aliquots were removed, diluted in ice cold LB broth and titered

35 for plaque-forming units on a lawn of XL1-Blue(TM) cells. Table 114 shows that incubation of the MK-BPTI phage with

84

immobilized anhydro-trypsin results in a very significant
loss in titer over a four hour period while no such
effect is observed with the MK (control) phage.   The
reduction in phage titer is also proportional to the
5   amount of immobilized anhydro-trypsin added to the MK-
BPTI phage.   Incubation with 5 μl of a 50% slurry of
agarose-immobilized streptavidin (Sigma, St. Louis, MO)
in TBS/BSA buffer does not reduce the titer of either the
MK-BPTI or MK phage.   These data are consistent with the
10   presentation of a correctly-folded, functional BPTI
protein on the surface of the MK-BPTI phage but not on
the MK phage.   Unfolded or incomplete BPTI domains are
not expected to bind anhydro-trypsin.   Furthermore,
unfolded BPTI domains are expected to be non-specifically
15   sticky.

NEUTRALIZATION OF PHAGE TITER WITH ANTI-BPTI ANTIBODY

MK-BPTI and MK phage were diluted to a concentration of
$4 \cdot 10^8$ plaque-forming units per ml in LB broth.   15 μl of
diluted phage were added to an equivalent volume of
20   either rabbit anti-BPTI serum or normal rabbit serum
(both diluted 10-fold in LB broth).   Following incubation
at 37°C, aliquots were removed, diluted by 10^4 in ice-cold
LB broth and titered for pfus on a lawn of XL1-Blue[TM].
Incubation of the MK-BPTI phage with anti-BPTI serum
25   results in a steady loss in titer over a two hour period
while no such effect is observed with the MK phage.   As
expected, normal rabbit serum does not reduce the titer
of either the MK-BPTI or the MK phage.   Prior incubation
of the anti-BPTI serum with authentic BPTI protein but
30   not with an equivalent amount of E. coli protein, blocks
the ability of the serum to reduce the titer of the MK-
BPTI phage.   These data are consistent with the
presentation of BPTI-specific pitopes on the surface of
the MK-BPTI phage but not the MK phage.   More specifi-
35   cally, the data indicates that these BPTI epitopes are
associated with the gene III protein and that association

85

of this fusion protein with an anti-BPTI antibody blocks
its ability to mediate the infection of cells.

NEUTRALIZATION OF PHAGE TITER WITH TRYPSIN

MK-BPTI and MK phage were diluted to a concentration of
5    $4 \cdot 10^8$ plaque-forming units per ml in LB broth. Diluted
phage were added to an equivalent volume of trypsin
diluted to various concentrations in LB broth. Following
incubation at 37°C, aliquots were removed, diluted by 10^4
in ice cold LB broth and titered for plaque-forming units
10   on a lawn of XL1-Blue™). Incubation of the MK-BPTI
phage with 0.15 μg of trypsin results in a 70% loss in
titer after two hours while only a 15% loss in titer is
observed for MK phage. A reduction in the amount of
trypsin added to phage results in a reduction in the loss
15   of titer. However, at all trypsin concentrations inves-
tigated, the MK-BPTI phage are more sensitive to incuba-
tion with trypsin than the MK phage. Thus, association
of the BPTI-III fusion protein displayed on the surface
of the MK-BPTI phage with trypsin blocks its ability to
20   mediate the infection of cells.

The reduction in titer of phage MK by trypsin is an
example of a phenomenon that is probably general:
proteases, if present in sufficient quantity, will
degrade proteins on the phage and reduce infectivity.
25   The present application lists several means that can be
used to overcome this problem.

AFFINITY SELECTION SYSTEM

Affinity Selection with Immobilized Anhydro-Trypsin

MK-BPTI and MK phage were diluted to a concentration of
30   $1.4 \cdot 10^{12}$ particles per ml in TBS buffer (PARM88)
containing 1.0 mg/ml BSA. We added $4.0 \cdot 10^{10}$ phage to 5 μl
of a 50% slurry of either agarose-immobilized anhydro-
trypsin beads (Pierce Chemical Co.) or agarose-
immobilized streptavidin beads (Sigma) in TBS/BSA.
35   Following a 3 hour incubation at room temperature, the
beads were pelleted by centrifugation for 30 seconds at

5000 rpm in a microfuge and the supernatant fraction was collected. The beads were washed 5 times with TBS/Tween buffer (PARM88) and after each wash the beads were pelleted by centrifugation and the supernatant was
5 removed. Finally, beads were resuspended in elution buffer (0.1 N HCl containing 1.0 mg/ml BSA adjusted to pH 2.2 with glycine) and following a 5 minute incubation at room temperature, the beads were pelleted by centrifugation. The supernatant was removed and neutralized by the
10 addition of 1.0 M Tris-HCl buffer, pH 8.0.

Aliquots of phage samples were applied to a Nytran membrane using a Schleicher and Schuell (Keene, NH) filtration minifold and phage DNA was immobilized onto the Nytran by baking at 80°C for 2 hours. The baked
15 filter was incubated at 42°C for 1 hour in pre-wash solution (MANI82) and pre-hybridization solution (5Prime-3Prime, West Chester, PA). The 1.0 Kb NarI (base 1630)/XmnI (base 2646) DNA fragment from MK RF was radioactively labelled with ^{32}P-dCTP using an oligolabell-
20 ing kit (Pharmacia, Piscataway, NJ). The radioactive probe was added to the Nytran filter in hybridization solution (5Prime-3Prime) and, following overnight incubation at 42°C, the filter was washed and autoradiographed.

The efficiency of this affinity selection system can be
25 semi-quantitatively determined using a dot-blot procedure. Exposure of MK-BPTI-phage-treated anhydro-trypsin beads to elution buffer releases bound MK-BPTI phage. Streptavidin beads do not retain phage MK-BPTI. Anhydro-trypsin beads do not retain phage MK. In the
30 experiment depicted in Table 115, we estimate that 20% of the total MK-BPTI phage were bound to 5 μl of the immobilized anhydro-trypsin and were subsequently recovered by washing the beads with elution buffer (pH 2.2 HCl/glycine). Under the same conditions, no
35 detectable MK-BPTI phage were bound and subsequently recovered from the streptavidin beads. The amount of MK-

87

BPTI phage recovered in the elution fraction is
proportional to the amount of immobilized anhydro-trypsin
added to the phage. No detectable MK phage were bound to
either the immobilized anhydro-trypsin or streptavidin
5   beads and no phage were recovered with elution buffer.
These data indicate that the affinity selection system
described above can be utilized to select for phage
displaying a specific folded protein (in this case,
BPTI). Unfolded or incomplete BPTI domains are not
10   expected to bind anhydro-trypsin.

Affinity Selection with Anti-BPTI antibodies

MK-BPTI and MK phage were diluted to a concentration of
$1 \cdot 10^{10}$ particles per ml in Tris buffered saline solution
(PARM88) containing 1.0 mg/ml BSA. Two·10^8 phage were
15   added to 2.5 μg of either biotinylated rabbit anti-BPTI
IgG in TBS/BSA or biotinylated rabbit anti-mouse antibody
IgG (Sigma) in TBS/BSA, and incubated overnight at 4°C.
A 50% slurry of streptavidin-agarose (Sigma), washed
three times with TBS buffer prior to incubation with 30
20   mg/ml BSA in TBS buffer for 60 minutes at room tempera-
ture, was washed three times with TBS/Tween buffer
(PARM88) and resuspended to a final concentration of 50%
in this buffer. Samples containing phage and
biotinylated IgG were diluted with TBS/Tween prior to the
25   addition of streptavidin-agarose in TBS/Tween buffer.
Following a 60 minute incubation at room temperature,
streptavidin-agarose beads were pelleted by centrifu-
gation for 30 seconds and the supernatant fraction was
collected. The beads were washed 5 times with TBS/Tween
30   buffer and after each wash, the beads were pelleted by
centrifugation and the supernatant was removed. Finally,
the streptavidin-agarose beads were resuspended in
elution buffer (0.1 N HCl containing 1.0 mg/ml BSA
adjusted to pH 2.2 with glycine), incubated 5 minute at
35   room temperature, and pelleted by centrifugation. The

88

supernatant was removed and neutralized by the addition
of 1.0 M Tris-HCl buffer, pH 8.0.

   Aliquots of phage samples were applied to a Nytran
membrane using a Schleicker and Schuell minifold appar-
5   atus.   Phage DNA was immobilized onto the Nytran by
baking at 80°C for 2 hours. Filters were washed for 60
minutes in pre-wash solution (MANI82) at 42°C then
incubated at 42°C for 60 minutes in Southern pre-hybri-
dization solution (5Prime-3Prime).   The 1.0 Kb NarI
10 (1630bp)/XmnI (2646 bp) DNA fragment from MK RF was
radioactively labelled with ^{32}P-αdCTP using an oligo-
labelling kit (Pharmacia, Piscataway, NJ).   Nytran
membranes   were   transferred   from   pre-hybridization
solution to Southern hybridization solution (5Prime--
15 3Prime) at 42°C. The radioactive probe was added to the
hybridization solution and following overnight incubation
at 42°C, the filter was washed 3 times with 2 x SSC, 0.1%
SDS at room temperature and once at 65°C in 2 x SSC, 0.1%
SDS. Nytran membranes were subjected to autoradiography.
20 The efficiency of the affinity selection system can be
semi-quantitatively determined using the above dot blot
procedure.   Comparison of dots A1 and B1 or C1 and D1
indicates that the majority of phage did not stick to the
streptavidin-agarose beads.   Washing with TBS/Tween
25 buffer removes the majority of phage which are non-
-specifically   associated   with   streptavidin   beads.
Exposure of the streptavidin beads to elution buffer
releases bound phage only in the case of MK-BPTI phage
which have previously been incubated with biotinylated
30 rabbit anti-BPTI IgG.   This data indicates that the
affinity selection system described above can be utilized
to select for phage displaying a specific antigen (in
this case BPTI). We estimate an enrichment factor of at
least 40 fold based on the calculation
35
                              Percent MK-BPTI phage recovered
   Enrichment Factor = ─────────────────────────────────

89

Percent MK phage recovered

EXAMPLE III

5   **BPTI:VIII BOUNDARY EXTENSIONS.**

To increase the flexibility between BPTI and mature
gVIIIp, we introduced codons for peptide extensions
between these domains.

Adding block extensions to the fusion protein interface.

10 The M13 gene III product contains 'stalk-like' regions
as implied by electron micrographic visualization of the
bacteriophage (LOPE85). The predicted amino acid
sequence of this protein contains repeating motifs, which
include:

15 glu.gly.gly.gly.ser (EGGGS) seven times
gly.gly.gly.ser (GGGS) three times
glu.gly.gly.gly.thr (EGGGT) once.

The aim of this section was to insert, at the domain
interface, multiple unit extensions which would mirror

20 the repeating motifs observed in the III gene product.

Two synthetic oligonucleotides were synthesized. We
picked the third base of these codons so that translation
of the oligonucleotide in the opposite direction would
yield SER. When annealed the synthetic oligonucleotides

25 give the following unit duplex sequence (an EGGGS
linker):

```
            E   G   G   G   S
    5' C.GAG.GGA.GGA.GGA.TC     3'
    3'   TC.CCT.CCT.CCT.AGG.C 5'
            (L) (S) (S) (S) (G)
```
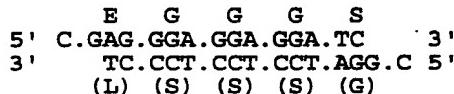
30 The duplex has a common two base pair 5' overhang (GC)
at either end of the linker which allows for both the
ligation of multiple units and the ability to clone into
the unique NarI recognition sequence present in OCV's

35 M13MB48 and Gem MB42. This site is positioned within 1
codon of the DNA encoding the interface. The cloning of
an EGGGS linker (or multiple linker) into the vector NarI

90

site destroys this recognition sequence. Insertion of
the EGGGS linker in reverse orientation leads to
insertion of GSSSL into the fusion protein.

Addition of a single EGGGS linker at the NarI site of
5   the gene shown in Table 113 leads to the following gene:

```
            79   80  80a 80b 80c 80d 80e 81  82  83  84
             G    G   E   G   G   G   S   A   A   E   G
                      ----------------
10                   GGT.GGC.GAG.GGA.GGA.GGA.TCC.GCC.GCT.GAA.GGT
                      ------------------
```

Note that there is no preselection for the orientation
of the linker(s) inserted into the OCV and that multiple
15 linkers of either orientation (with the predicted EGGGS
or GSSSL amino acid sequence) or a mixture of
orientations (inverted repeats of DNA) could occur.

A ladder of increasingly large multiple linkers was
established by annealing and ligating the two starting
20 oligonucleotides containing different proportions of 5'
phosphorylated and non-phosphorylated ends. The logic
behind this is that ligation proceeds from the 3' unphos-
phorylated end of an oligonucleotide to the 5' phosphor-
ylated end of another. The use of a mixture of phosphor-
25 ylated and non-phosphorylated oligonucleotides allows for
an element of control over the extent of multiple linker
formation. A ladder showing a range of insert sizes was
readily detected by agarose gel electrophoresis spanning
15 bp (1 unit duplex-5 amino acids) to greater than 600
30 base pairs (40 ligated linkers-200 amino acids).

Large inverted repeats can lead to genetic instability.
Thus we chose to remove them, prior to ligation into the
OCV, by digesting the population of multiple linkers with
the restriction enzymes AccIII or XhoI, since the
35 linkers, when ligated 'head-to-head' or 'tail-to-tail',
generate these recognition sequences. Such a digestion
significantly reduces the range in sizes of the multiple
linkers to between 1 and 8 linker units (i.e. between 5

91

and 40 amino acids in steps of 5), as assessed by agarose
gel electrophoresis.

The linkers were ligated (as a pool of different insert
sizes or as gel-purified discrete fragments) into NarI
5   cleaved OCVs M13MB48 or GemMB42 using standard methods.
Following ligation the restriction enzyme NarI was added
to remove the self-ligating starting OCV (since linker
insertion destroys the NarI recognition sequence). This
mixture was used to transform XL-1 blue cells and appro-
10   priately plated for plaques (OCV M13MB48) or ampicillin
resistant colonies (OCV GemMB42).

The transformants were screened using dot blot DNA
analysis with one of two ^{32}P labeled oligonucleotide
probes. One probe consisted of a sequence complementary
15   to the DNA encoding the P1 loop of BPTI while the second
had a sequence complementary to the DNA encoding the
domain interface region. Suitable linker candidates
would probe positively with the first probe and
negatively or poorly with the second. Plaque purified
20   clones were used to generate phage stocks for binding
analyses and BPTI display while the Rf DNA derived from
phage infected cells was used for restriction enzyme
analysis and sequencing. Representative insert sequences
of selected clones analyzed are as follows:
25
M13.3X4   (GG)C.GGA.TCC.TCC.TCC.CT(C.GCC)
                         gly ser ser ser leu

M13.3X7   (G C.GAG.GGA.GGA.GGA.TC(C.GCC)
30   glu gly gly gly ser

M13.3X11   (GG)C.GAG.GGA.GGA.GGA.TCC.GGA.TCC.TCC.
                         glu gly gly gly ser gly ser ser

35   TCC.CTC.GGA.TCC.TCC.TCC.CT(C.GCCC)
                         ser leu gly ser ser ser leu

These highly flexible oligomeric linkers are believed to
be useful in joining a binding domain to the major coat
40   (gene VIII) protein of filamentous phage to facilitate

92

the display of the binding domain on the phage surface.
They may also be useful in the construction of chimeric
OSPs for other genetic packages as well.

5   **Incorporation of interdomain extension fusion proteins
       into phage.**
            A phage pool containing a variegated pentapeptide
       extension at the BPTI:coat protein interface was used to
       infect SEF' cells. Using the criteria of the previous
10   section, we determined that extended fusion proteins were
       incorporated into phage. Gel electrophoresis of the
       generated phage, followed by silver staining or western
       analysis with anti-BPTI rabbit serum, demonstrated fusion
       proteins that migrated similarly to, but discernably
15   slower than, the starting fusion protein.
            With regard to the 'EGGGS linker' extensions of the
       domain interface, individual phage stocks predicted to
       contain one or more 5-amino-acid unit extensions were
       analyzed in a similar fashion. The migration of the
20   extended fusion proteins were readily distinguishable
       from the parent fusion protein when viewed by western
       analysis or silver staining. Those clones analyzed in
       more detail included M13.3X4 (which contains a single
       inverted EGGGS linker with a predicted amino acid
25   sequence of GSSSL), M13.3X7 (which contains a correctly
       orientated linker with a predicted amino acid sequence of
       EGGGS), M13.3X11 (which contains 3 linkers with an
       inversion and a predicted amino acid sequence for the
       extension of EGGGSGSSSLGSSSL) and M13.3Xd which contains
30   an extension consisting of at least 5 linkers or 25 amino
       acids.
            The extended fusion proteins were all incorporated into
       phage at high levels (on average 10's of copies per phage
       were present and when analyzed by gel electrophoresis
35   migrated rates consistent with the predicted size of the
       extension. Clones M13.3X4 and M13.3X7 migrated at a

93

position very similar to but discernably different from
the parent fusion protein, while M13.3X11 and M13.3Xd
w e r e   m a r k e d l y   l a r g e r .

5
                              EXAMPLE IV
                             Peptide phage

      The following materials and methods were used in the
10   examples which follow.
      1. Peptide Phage
      HPQ6, a putative disulfide-bonded mini-protein, was
displayed on M13 phage as an insert in the gene III
protein (gIIIp). M13 has about five copies of gIIIp per
15   virion. The phage were constructed by standard methods.
HPQ6 includes the sequence CHPQFPRC characteristic of
Devlin's streptavidin-binding E peptide (DEVL90), as well
as a F.X$_a$ recognition site (see Table 820). HPQ6 phage
were shown to bind to streptavidin.
20   An unrelated display phage with no affinity for
streptavidin, MKTN, was used as a control.
      2. Streptavidin.
      Commercially available immobilized to agarose beads
(Pierce). Streptavidin (StrAv) immobilized to 6% beaded
25   agarose at a concentration of 1 to 2 mg per ml gel,
provided as a 50% slurry. Also available as free protein
(Pierce) with a specific activity of 14.6 units per mg (1
unit will bind 1 μg of biotin). A stock solution of 1 mg
per ml in PBS containing 0.01% azide is made.

30   3. D-Biotin.

      Commercially available (Boehringer Mannheim) in
crystallized form. A stock solution of 4 mM is made.

      4. Streptavidin coating of microtiter well plates.

      Immulon (#2 or #4) strips or plates are used. 100μL of
35   StrAv stock is added to each 250 μL capacity well and
incubated overnight at 4°C. The stock is removed and

94

replaced with 250 μL of PBS containing BSA at a concentration of 1 mg per mL and left at 4°C for a further 1 hour. Prior to use in a phage binding assay the wells are washed rapidly 5 times with 250 μL of PBS

5   containing 0.1% Tween.

5. Binding Assays.

Bead Assay.

Between 10 and 20 μL of the StrAv bead slurry (5 to 10 μL bead volume) is washed 3 times with binding buffer

10   (TBS containing BSA at a concentration of 1 mg per mL) just prior to the binding assay. 50 to 100 μL of binding buffer containing control or peptide-display phage (10^8 to 10^{11} total plaque forming units - pfu's) is added to each microtube. Binding is allowed to proceed for 1 hour

15   at room temperature using an end over end rotator. The beads are briefly centrifuged and the supernatant removed. The beads are washed a further 5 times with 1 mL of TBS containing 0.1% Tween, each wash consisting of a 5 min incubation and a brief centrifugation. Finally

20   the bound phage are eluted from the StrAv beads by a 10 min incubation with pH 2 citrate buffer containing 1 mg per mL BSA which is subsequently neutralized with 260 μL of 1M tris pH 8. The number of phage present in each step is determined as plaque forming units (pfu's)

25   following appropriate dilutions and plating in a lawn of F' containing *E. coli*.

Plate Assay.

To each StrAv-coated well is added 100 μL of binding buffer (PBS with 1 mg per mL BSA) containing a known

30   quantity of phage (between 10^8 and 10^{11} pfu's). Incubation proceeds for 1 hr at room temperature followed by removal of the non-bound phage and 10 rapid washes with PBS 0.1% Tween. The bound phage are eluted with 250 μL of pH2 citrate buffer containing 1 mg per mL BSA and

95

neutralization with 60 μL of 1M tris pH 8. The number of
phage present in each step is determined as plaque
forming units (pfu's) following appropriate dilutions and
plating in a lawn of F' containing *E. coli*.

5   **EXAMPLE V**
                    Effect of Dithiothreitol (DTT)
             on display phage binding to streptavidin-agarose.

<u>Preliminary control experiments.</u>

a. Use of HRP-conjugated biotin and streptavidin beads.
10   Binding capacity of StrAv agarose beads for HRP-
conjugated biotin determined to be ~ 1 μg (equivalent to
~ 150 pmol biotin) per 5 μL beads (the amount used in
these experiments).

b. Effect of DTT on HRP-conjugated biotin binding to
15   StrAv beads.
      5 μL of StrAv beads were incubated with 10 ng of HRP-
biotin in binding buffer (TBS-BSA) in the presence of
varying amounts of DTT (at least 99% reduced). Following
a 15 minute incubation at room temperature, the beads
20   were washed two times in binding buffer and an HRP
substrate added. Color development was allowed to
proceed and noted in a semi-quantitative manner. Table
827 shows that the binding of biotinylated horseradish
peroxidase (HRP) is not greatly affected by
25   concentrations of DTT below 20 mM. *N.B.* DTT
concentrations of 20 and 50 mM also inhibited the
interaction of HRP and substrate in the absence of StrAv
beads hence having a general negative effect in this
system.

96

Effect of DTT on HPQ6 display phage infectivity.

10⁸ pfus of HPQ6 were added to binding buffer (TBS-BSA)
in the presence of different concentrations of DTT.
Incubated at room temperature for 1 hour then diluted and
5   plated to determine titer as pfus. Table 828 show the
effect of DTT on the infectivity of phage HPQ6. Hence,
either DTT has no effect on phage infectivities over this
range of concentrations or the effects are reversed on
dilution of the phage. From these control experiments it
10   is apparent that DTT can be used at concentrations below
10 mM in studies on the effect of reducing agents on
peptide display phage binding to StrAv.

Table 829 shows the effect of DTT on the binding of
phage HPQ6 and MKTN to StrAv beads. The most significant
15   effect of DTT on HPQ6 binding to StrAv occurred between
0.1 and 1.0 mM DTT, a concentration at which no negative
effects were observed in the preliminary control
experiments. These results strongly indicate that, in
the case of HPQ6 display phage, DTT has a marked effect
20   on binding to StrAv and that the presence of a disulfide
bridge within the displayed peptide is a requirement for
good binding.

<div align="center">

EXAMPLE VI
RELEASE OF STREPTAVIDIN-BOUND DISPLAY PHAGE
BY FACTOR Xa CLEAVAGE

</div>

25   Phage HPQ6 contains a bovine F.X_a recognition site
(Y**IEGR**/IV). In many instances, IEGR is sufficient
recognition site for F.X_a, but we have extended the site
in each direction to facilitate efficient cleavage. The
30   effect of preincubating HPQ6 phage with F.X_a on binding to
StrAv beads is shown in Table 832. Thus while this
concentration of F.X_a (2.5 units) had no measurable effect
on the titer of the treated display phage it had a very
marked effect on the ability of the treated display phage
35   to bind to StrAv. This is consistent with the StrAv

97

recognition sequence being removed by the action of FX_a
recognizing and cleaving the YIEGR/IV sequence.

Table 833 shows the effect of FX_a treatment of HPQ6
following binding to StrAv. Is it possible to remove
5   display phage bound to their target by the use of FX_a in
place of pH or chaotropic agent elution? HPQ6 display
phage were allowed to bind to StrAv then incubated either
in FX_a buffer or the same buffer together with 2.5 units
of FX_a for 3 hrs. The amount eluted was compared to the
10   total number of phage bound as judged by a pH2 elution.
Therefore, while the display phage are slowly removed in
the buffer alone, the presence of FX_a significantly
increases this rate.

The removal of HPQ6 display phage from StrAv by FX_a was
15   also studied as a function of the amount of enzyme added
and the time of incubation, as shown in Table 834. *N.B.*
at greater concentrations of the enzyme (1.2 U for 1 hour
or 2.5 U for 2 hours), a loss in infectivity of the
treated phage was noted as measured by pfus.

20

98

Table 10: Abundances obtained
from various vgCodons

A. Optimized fxS Codon, Restrained by [D]+[E] = [K]+[R]

5

	T	C	A	G	
1 |	.26	.18	.26	.30	f
2 |	.22	.16	.40	.22	x
3 |	.5	.0	.0	.5	S

10

Amino acid	Abundance	Amino acid	Abundance
A	4.80%	C	2.86%
D	6.00%	E	6.00%
F	2.86%	G	6.60%
H	3.60%	I	2.86%
K	5.20%	L	6.82%
M	2.86%	N	5.20%
P	2.88%	Q	3.60%
R	6.82%	S	7.02% mfaa
T	4.16%	V	6.60%
W	2.86% lfaa	Y	5.20%
stop	5.20%		

15

20

99

$$[D] + [E] = [K] + [R] = .12$$

$$\text{ratio} = \text{Abun}(W)/\text{Abun}(S) = 0.4074$$

j	$(1/\text{ratio})^j$	$(\text{ratio})^j$	stop-free
1	2.454	.4074	.9480
2	6.025	.1660	.8987
3	14.788	.0676	.8520
4	36.298	.0275	.8077
5	89.095	.0112	.7657
6	218.7	$4.57 \cdot 10^{-3}$	.7258
7	536.8	$1.86 \cdot 10^{-3}$	.6881

100

Table 10: Abundances obtained

from various vgCodon

(continued)

5   B. Unrestrained, optimized

	T	C	A	G
1	.27	.19	.27	.27
2	.21	.15	.43	.21
3	.5	.0	.0	.5

Amino acid	Abundance	Amino acid	Abundance
A	4.05%	C	2.84%
D	5.81%	E	5.81%
F	2.84%	G	5.67%
H	4.08%	I	2.84%
K	5.81%	L	6.83%
M	2.84%	N	5.81%
P	2.85%	Q	4.08%
R	6.83%	S	6.89% mfaa
T	4.05%	V	5.67%
W	2.84% lfaa	Y	5.81%
stop	5.81%		

25

101

$[D] + [E] = 0.1162 \quad [K] + [R] = 0.1264$

ratio = Abun(W)/Abun(S) = 0.41176

5

j	$(1/ratio)^j$	$(ratio)^j$	stop-free
1	2.4286	.41176	.9419
2	5.8981	.16955	.8872
3	14.3241	.06981	.8356
4	34.7875	.02875	.7871
5	84.4849	.011836	.74135
6	205.180	.004874	.69828
7	498.3	$2.007 \cdot 10^{-3}$	.6577

10

15

102

Table 10: Abundances obtained

from various vgCodon

(continued)

5   C. Optimized NNT

	T	C	A	G
1	.2071	.2929	.2071	.2929
2	.2929	.2071	.2929	.2071
3	1.	.0	.0	.0

Amino acid	Abundance	Amino acid	Abundance
A	6.06%	C	4.29% lfaa
D	8.58%	E	none
F	6.06%	G	6.06%
H	8.58%	I	6.06%
K	none	L	8.58%
M	none	N	6.06%
P	6.06%	Q	none
R	6.06%	S	8.58% mfaa
T	4.29% lfaa	V	8.58%
W	none	Y	6.06%
stop	none		

103

j	$(1/ratio)^j$	$(ratio)^j$	stop-free
1	2.0	.5	1.
2	4.0	.25	1.
3	8.0	.125	1.
4	16.0	.0625	1.
5	32.0	.03125	1.
6	64.0	.015625	1.
7	128.0	.0078125	1.

104

Table 10: Abundances obtained

from various vgCodon

(continued)

5

D. Optimized NNG

	T	C	A	G
1	.23	.21	.23	.33
2	.215	.285	.285	.215
3	.0	.0	.0	1.0

Amino acid	Abundance	Amino acid	Abundance
A	9.40%	C	none
D	none	E	9.40%
F	none	G	7.10%
H	none	I	none
K	6.60%	L	9.50% mfaa
M	4.90%	N	none
P	6.00%	Q	6.00%
R	9.50%	S	6.60%
T	6.6 %	V	7.10%
W	4.90% lfaa	Y	none
stop	6.60%		

105

$\underline{j}$	$\underline{(1/ratio)^j}$	$\underline{(ratio)^j}$	$\underline{stop\text{-}free}$
1	1.9388	.51579	0.934
2	3.7588	.26604	0.8723
3	7.2876	.13722	0.8148
4	14.1289	.07078	0.7610
5	27.3929	$3.65 \cdot 10^{-2}$	0.7108
6	53.109	$1.88 \cdot 10^{-2}$	0.6639
7	102.96	$9.72 \cdot 10^{-3}$	0.6200

106

Table 10: Abundances obtained

from optimum vgCodon

(continued)

5

E. Unoptimized NNS (NNK gives identical distribution)

	T	C	A	G
1	.25	.25	.25	.25
2	.25	.25	.25	.25
3	.0	.5	.0	0.5

15

Amino acid	Abundance	Amino acid	Abundance
A	6.25%	C	3.125%
D	3.125%	E	3.125%
F	3.125%	G	6.25%
H	3.125%	I	3.125%
K	3.125%	L	9.375%
M	3.125%	N	3.125%
P	6.25%	Q	3.125%
R	9.375%	S	9.375%
T	6.25%	V	6.25%
W	3.125%	Y	3.125%

107

stop   3.125%

5	j	$(1/\text{ratio})^j$	$(\text{ratio})^j$	stop-free
	1	3.0	.33333	.96875
	2	9.0	.11111	.9385
	3	27.0	.03704	.90915
	4	81.0	.01234567	.8807
10	5	243.0	.0041152	.8532
	6	729.0	$1.37 \cdot 10^{-3}$	.82655
	7	2187.0	$4.57 \cdot 10^{-4}$	.8007

108

Table 102b : Annotated Sequence of gene

after insertion of SalI linker

5   nucleotide

                                                        number

     5'-(GGATCC TCTAGA GTC) GGC-   3

        from pGEM polylinker

10

     tttaca CTTTATGCTTCCGGCTCG tataat GTGTGG-   39

        -35   lacUV5   -10

15

        aATTGTGAGCGcTcACAATT-   59

        lacO-symm operator

20   gagctc   AGAGG   CttaCT-   77

        SacI   Shine-Dalgarno seq.

25 |fM | K | K | S | L | V | L | K | A | S |

      | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10|

      |ATG|AAG|AAA|TCT|CTG|GTT|CTT|AAG|GCT|AGC|-   107

109

```
                        | Afl II| Nhe I |


        | V | A | V | A | T | L | V | P | M | L |
   5    | 11| 12| 13| 14| 15| 16| 17| 18| 19| 20|
        |GTT|GCT|GTC|GCG|ACC|CTG|GTA|CCT|ATG|TTG|-     137
            | Nru I|        | Kpn I|



   10   | S | F | A | R | P | D | F | C | L | E |
        | 21| 22| 23| 24| 25| 26| 27| 28| 29| 30|
        |TCC|TTC|GCT|CGT|CCG|GAT|TTC|TGT|CTC|GAG|-     167
              ↑    |AccIII|              | Ava I |
         M13/BPTI Jnct                   | Xho I |

   15


        | P | P | Y | T | G | P | C | K | A | R |
        | 31| 32| 33| 34| 35| 36| 37| 38| 39| 40|
        |CCA|CCA|TAC|ACT|GGG|CCC|TGC|AAA|GCG|CGC|_     197
   20       | PflM I      |     ||      |BssH II|
                   | Apa I ||
                   | Dra II |
                   | Pss I  |
```

110

Table 102b : Annotated Sequence

of gene after insertion of SalI linker

(continued)

5

```
| I | I | R | Y | F | Y | N | A | K | A |
| 41| 42| 43| 44| 45| 46| 47| 48| 49| 50|
|ATC|ATC|CGC|TAT|TTC|TAC|AAT|GCT|AAA|GC |-     226
```

10

```
| G | L | C | Q | T | F | V | Y | G | G |
| 51| 52| 53| 54| 55| 56| 57| 58| 59| 60|
A|GGC|CTG|TGC|CAG|ACC|TTT|GTA|TAC|GGT|GGT|-     257
| Stu I|                          | Acc I |
```

15
```
                                   | Xca I |
```


```
 | C | R | A | K | R | N | N | F | K |
 | 61| 62| 63| 64| 65| 66| 67| 68| 69|
20|TGC|CGT|GCT|AAG|CGT|AAC|AAC|TTT|AAA|-     284
         | Esp I   |
```


```
 | S | A | E | D | C | M | R | T | C | G |
 | 70| 71| 72| 73| 74| 75| 76| 77| 78| 79|
25|TCG|GCC|GAA|GAT|TGC|ATG|CGT|ACC|TGC|GGT|-     314
   |XmaIII|          | Sph I|
```

111

BPTI/M13 boundary

ᵥ|

| G | A | A | E | G | D | D | P | A | K | A | A |
| 80| 81| 82| 83| 84| 85| 86| 87| 88| 89| 90| 91|

5 |GGC|GCC|GCT|GAA|GGT|GAT|GAT|CCG|GCC|AAG|GCG|GCC| - 350

| Bbe I | | Sfi I |

| Nar I |

| F | N | S | L | Q | A | S | A | T |
| 92| 93| 94| 95| 96| 97| 98| 99|100|

10 |TTC|AAT|TCT|CTG|CAA|GCT|TCT|GCT|ACC| -   377

                      |Hind 3|

15

| E | Y | I | G | Y | A | W |
|101|102|103|104|105|106|107|

|GAG|TAT|ATT|GGT|TAC|GCG|TGG| -   398

20 | A | M | V | V | V | I | V | G | A |
|108|109|110|111|112|113|114|115|116|

|GCC|ATG|GTG|GTG|GTT|ATC|GTT|GGT|GCT| -   425

| BstX I |

| Nco I|

112

Table 102b: Annotated Sequence
after insertion of <u>Sal</u>I linker
(continued)

5

```
    | T | I | G | I |
    |117|118|119|120|
    |ACC|ATC|GGG|ATC|-                      437
```

10

```
    | K | L | F | K | K | F | T | S | K | A |
    |121|122|123|124|125|126|127|128|129|130|
    |AAA|CTG|TTC|AAG|AAG|TTT|ACT|TCG|AAG|GCG|-    467
                                    |Asu II|
```

15

```
    | S | . | . | . |
    |131|132|133|134|
    ·|TCT|TAA|TGA|TAG|    GGTTACC-                486
                         BstE II
```

20

```
AGTCTA AGCCCGC CTAATGA GCGGGCT TTTTTTTT-         521
    terminator
```

25

```
    aTCGA    GACctgca GGTCGACC ggcatgc-3'
                |SalI|
```

Note the following enzyme equivalences,

Xma III   = Eag I   Acc III   = BspM II

Dra II   = EcoO109 I   Asu II   = BstB I

5

-114-

Table 106: Signal Peptides

PhoA	M K q s t i a l a l l p l l f t p v t K A /R T... (17)
MalE	M K I K T G A R i l a l s a l t t m m f s a s a l a /K I... (18)
OmpF	M M K R n i l a v i v p a l l v a g t a n a /a K... (19)
Bla	M S I Q H F R v a l i p f f a a f c l p v f a /h p... (>18)
LamB	M M I T L R K l p l a v a v a a g v m s a q a m a /v D... (19)
Lpp	M K A T K l v l g a v i l g s t l l a g /c s... (>17)
gpIII	M K K l l f a i p l v v p f y s h s /a K T V K... (16)
gpIII-BPTI	M K K l l f a i p l v v p f y s g a /R P D... (15)
gpVIII	M K K S L V L K a s v a v a t l v p m l s f a /a K G D D... (16)
gpVIII-BPTI	M K K S L V L K a s v a v a t l v p m l s f a /R P D... (15)
gpVIII'	M K K s l v l l a s v a v a t l v p m l s f a /a K G D D... (21)

5

10

15

115

Table 107: In vitro transcription/translation
analysis of vector-encoded
signal::BPTI::mature VIII protein species

	31 kd species[a]	14.5 kd species[b]
No DNA (control)	-[c]	-
pGEN-3Zf(-)	+	-
pGEM-MB16	+	-
pGEM-MB20	+	+
pGEM-MB26	+	+
pGEM-MB42	+	+
pGEM-MB46	ND	ND

Notes:

a.) pre-beta-lactamase, encoded by the amp (bla) gene.

b.) pre-BPTI/VIII peptides encoded by the synthetic gene and derived constructs.

c.) - for absence of product; + for presence of product; ND for Not Determined.

116

Table 108: Western analysis[a] of _in vivo_
expressed
signal::BPTI::mature VIII protein species

5   A) expression in strain XL1-Blue

	signal	14.5 kd species[b]	12 kd species[c]
pGEM-3Zf(-)	-	-[d]	-
pGEM-MB16	VIII	-	-
pGEM-MB20	VIII	++	-
pGEM-MB26	VIII	+++	+/-
pGEM-MB42	phoA	++	+

10

B) expression in strain SEF'

	signal	14.5 kd species[b]	12 kd species[c]
pGEM-MB42	phoA	+/-	+++

15

Notes:

a) Analysis using rabbit anti-BPTI polyclonal antibodies
and horse-radish-peroxidase-conjugated goat anti-rabbit IgG
20   antibody.

b) pro-BPTI/VIII peptides encoded by the synthetic gene
and derived constructs.

c) processed BPTI/VIII peptide encoded by the synthetic
gene.

25   d) not present .... -
weakly present ..... +/-
present .... +
strong presence .... ++
very strong presence +++

Table 109: M13 gene III

```
      1579          5'-GT GAAAAAATTA TTATTCGCAA TTCCTTTAGT
      1611  TGTTCCTTTC TATTCTCACT CCGCTGAAAC TGTTGAAAGT
      1651  TGTTTAGCAA AACCCCATAC AGAAAATTCA TTTACTAACG
   5  1691  TCTGGAAAGA CGACAAAACT TTAGATCGTT ACGCTAACTA
      1731  TGAGGGTTGT CTGTGGAATG CTACAGGCGT TGTAGTTTGT
      1771  ACTGGTGACG AAACTCAGTG TTACGGTACA TGGGTTCCTA
      1811  TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA
      1851  GGGTGGCGGT TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT
   10 1891  ACTAAACCTC CTGAGTACGG TGATACACCT ATTCCGGGCT
      1931  ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCTGG
      1971  TACTGAGCAA AACCCCGCTA ATCCTAATCC TTCTCTTGAG
      2011  GAGTCTCAGC CTCTTAATAC TTTCATGTTT CAGAATAATA
      2051  GGTTCCGAAA TAGGCAGGGG GCATTAACTG TTTATACGGG
   15 2091  CACTGTTACT CAAGGCACTG ACCCCGTTAA AACTTATTAC
      2131  CAGTACACTC CTGTATCATC AAAAGCCATG TATGACGCTT
      2171  ACTGGAACGG TAAATTCAGA GACTGCGCTT TCCATTCTGG
      2211  CTTTAATGAG GATCCATTCG TTTGTGAATA TCAAGGCCAA
      2251  TCGTCTGACC TGCCTCAACC TCCTGTCAAT GCTGGCGGCG
   20 2291  GCTCTGGTGG TGGTTCTGGT GGCGGCTCTG AGGGTGGTGG
      2331  CTCTGAGGGT GGCGGTTCTG AGGGTGGCGG CTCTGAGGGA
      2371  GGCGGTTCCG GTGGTGGCTC TGGTTCCGGT GATTTTGATT
      2411  ATGAAAAGAT GGCAAACGCT AATAAGGGGG CTATGACCGA
      2451  AAATGCCGAT GAAAACGCGC TACAGTCTGA CGCTAAAGGC
   25 2491  AAACTTGATT CTGTCGCTAC TGATTACGGT GCTGCTATCG
      2531  ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA
      2571  TGGTGCTACT GGTGATTTTG CTGGCTCTAA TTCCCAAATG
      2611  GCTCAAGTCG GTGACGGTGA TAATTCACCT TTAATGAATA
      2651  ATTTCCGTCA ATATTTACCT TCCCTCCCTC AATCGGTTGA
   30 2691  ATGTCGCCCT TTTGTCTTTA GCGCTGGTAA ACCATATGAA
      2731  TTTTCTATTG ATTGTGACAA AATAAACTTA TTCCGTGGTG
      2771  TCTTTGCGTT TCTTTTATAT GTTGCCACCT TTATGTATGT
      2811  ATTTTCTACG TTTGCTAACA TACTGCGTAA TAAGGAGTCT
      2851  TAATCATGCC AGTTCTTTTG GGTATTCCGT
```

118

Table 110: Introduction of NarI into gene III

A) Wild-type III, portion encoding the signal peptide

```
5                     M   K   K   L   L   F   A   I   P   L
                      1   2   3   4   5   6   7   8   9   10
      1579      5'-GTG AAA AAA TTA TTA TTC GCA ATT CCT TTA
```

```
10                                       / Cleavage site
                                       v|
                V   V   P   F   Y   S   H   S   A   E   T   V
               11  12  13  14  15  16  17  18  19  20  21  22
      1609     GTT GTT CCT TTC TAT TCT CAC TCC GCT GAA ACT GTT-3'
15
```

B) III, portion encoding the signal peptide with NarI site

```
20                    m   k   k   l   l   f   a   I   p   l
                      1   2   3   4   5   6   7   8   9   10
      1579      5'-gtg aaa aaa tta tta ttc gca att cct tta
```

```
25                                       / cleavage site
                                       v|
                v   v   p   f   y   s   G   A   a   e   t   v
               11  12  13  14  15  16  17  18  19  20  21  22
      1609     gtt gtt cct ttc tat tct GGc Gcc gct gaa act gtt-3'
30
```

119

Table 113 : Annotated Sequence f
pGEM-MB42 comprising Ptac::RBS(GGAGGAAATAAA)::
<u>phoA-signal</u>::<u>mature-bpti</u>::<u>mature-VIII-coat-protein</u>

5

5'-GGATCC actccccatcccc
       └──┴──┘
       <u>BamHI</u>

10

ctg <u>TTGACA</u> attaatcatcg<u>GCTCG</u> <u>tataat</u> GTGTGG-
    -35   <u>tac</u>   -10

15

<u>aATTGTGAGCGcTcACAATT</u>-
 lacO-symm operator

| M | K | Q | S | T |
| 1 | 2 | 3 | 4 | 5 |

20

GAGCTCCATGGGAGAAAATAAA |ATG |AAA |CAA |AGC |ACG | -
<u>|SacI|</u>   <----- phoA signal peptide

25

| I | A | L | L | P | L | L | F | T | P | V | T |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|ATC |GCA |CTC |TTA |CCG |TTA |CTG |TTT |ACC |CCT |GTG |ACA | -
--------------- phoA signal continues ----------

30

(There are no residues 20-23.)

| K | A | R | P | D | F | C | L | E |
| 18 | 19 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|AAA |GCC |CGT |CCG |GAT |TTC |TGT |CTC |GAG | -

35

phoA signal->^| |<u>AccIII</u>| |<u>Ava I</u>|
phoA/BPTI Jnct |<u>Xho I</u>|
                 |<----- BPTI insert ---------

40

| P | P | Y | T | G | P | C | K | A | R |
| 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
|CCA |CCA |TAC |ACT |GGG |CCC |TGC |AAA |GCG |CGC | -

         |<u> PflM I |</u> || |<u>BssH II|</u>

45

                       |<u>Apa I</u>||
                       |<u>Dra II</u>|
                       |<u>Pss I</u>|

●

120

Table 113 : Annotated Sequence of
Ptac::RBS(GGAGGAAATAAA)::
phoA-signal::mature-bpti::mature-VIII-coat-protein gene
(continued)

5

```
  | I  | I  | R  | Y  | F  | Y  | N  | A  | K  | A  |
  | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
  |ATC |ATC |CGC |TAT |TTC |TAC |AAT |GCT |AAA |GC  | -
```

10

```
  | G  | L  | C  | Q  | T  | F  | V  | Y  | G  | G  |
  | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
A |GGC |CTG |TGC |CAG |ACC |TTT |GTA |TAC |GGT |GGT | -
  | Stu I|                        | Acc I    |
                                  | Xca I    |
```

15

20

```
  | C  | R  | A  | K  | R  | N  | N  | F  | K  |
  | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 |
  |TGC |CGT |GCT |AAG |CGT |AAC |AAC |TTT |AAA | -
       | Esp I    |
```

```
  | S  | A  | E  | D  | C  | M  | R  | T  | C  | G  |
  | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 |
  |TCG |GCC |GAA |GAT |TGC |ATG |CGT |ACC |TGC |GGT | -
  |XmaIII|              | Sph I|
-------------- BPTI insert----------------
```

25

BPTI/M13 boundary

30

```
  | G  | A  | A  | E  | G  | D  | D  | P  | A  | K  | A  | A  |
  | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 |
  |GGC |GCC |GCT |GAA |GGT |GAT |GAT |CCG |GCC |AAG |GCG |GCC | -
  | Bbe I|                      | Sfi I            |
  | Nar I|
-- BPTI-->|<----- mature gene VIII coat protein ----
```

35

40

```
  | F  | N  | S  | L  | Q  | A  | S  | A  | T  |
  | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |100 |
  |TTC |AAT |TCT |CTG |CAA |GCT |TCT |GCT |ACC | -
                 |Hind 3|
```

45

```
  | E  | Y  | I  | G  | Y  | A  | W  |
  |101 |102 |103 |104 |105 |106 |107 |
  |GAG |TAT |ATT |GGT |TAC |GCG |TGG | -
```

Table 113 : Annotated Sequence of
Ptac::RBS(GGAGGAAATAAA)::
phoA-signal::mature-bpti::mature-VIII-coat-protein gene
(continued)

5

A	M	V	V	V	I	V	G	A
108	109	110	111	112	113	114	115	116
GCC	ATG	GTG	GTG	GTT	ATC	GTT	GGT	GCT |-

10

|___ BstX I _____|
|_ Nco I_|

T	I	G	I
117	118	119	120
ACC	ATC	GGG	ATC |-

15

K	L	F	K	K	F	T	S	K	A
121	122	123	124	125	126	127	128	129	130
AAA	CTG	TTC	AAG	AAG	TTT	ACT	TCG	AAG	GCG |-

20

                             |Asu II|

S	.	.	.
131	132	133	134
TCT	TAA	TGA	TAG

25

            GGTTACC-
            BstE II

30

AGTCTA AGCCCGC CTAATGA GCGGGCT TTTTTTTT-
    terminator_____

aTCGA   GACctgca GGTCGAC-3'
35
              |SalI|

122

Table 114: Neutralization of Phage Titer Using
Agarose-immobilized Anhydro-Trypsin

		Percent Residual Titer As a Function of Time (hours)		
Phage Type	Addition	1	2	4
MK-BPTI	5 μl IS	99	104	105
	2 μl IAT	82	71	51
	5 μl IAT	57	40	27
	10 μl IAT	40	30	24
MK	5 μl IS	106	96	98
	2 μl IAT	97	103	95
	5 μl IAT	110	111	96
	10 μl IAT	99	93	106

Legend:

IS = Immobilized streptavidin

IAT = Immobilized anhydro-trypsin

123

Table 115: Affinity Selection of MK-BPTI Phage
on Immobilized Anhydro-Trypsin

Phage Type	Addition	Percent of Total Phage Recovered in Elution Buffer
MK-BPTI	5 μl IS	<<1[a]
	2 μl IAT	5
	5 μl IAT	20
	10 μl IAT	50
MK	5 μl IS	<<1[a]
	2 μl IAT	<<1
	5 μl IAT	<<1
	10 μl IAT	<<1

Legend:

IS = Immobilized streptavidin
IAT = Immobilized anhydro-trypsin
[a] not detectable.

124

Table 116: translation of <u>Signal-III</u>::<u>bpti</u>::<u>mature-III</u>

```
   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15
  fM    K    K    L    L    F    A    I    P    L    V    V    P    F    Y
 GTG  AAA  AAA  TTA  TTA  TTC  GCA  ATT  CCT  TTA  GTT  GTT  CCT  TTC  TAT
 |<------- gene III signal peptide ------------------------

  16   17   18   19   20   21   22   23   24   25   26   27   28   29   30
   S    G    A    R    P    D    F    C    L    E    P    P    Y    T    G
 TCT  GGC  GCC  cgt  ccg  gat  ttc  tgt  ctc  gag  cca  cca  tac  act  ggg
 ----------->|<----- BPTI insertion ----------------------

  31   32   33   34   35   36   37   38   39   40   41   42   43   44   45
   P    C    K    A    R    I    I    R    Y    F    Y    N    A    K    A
 ccc  tgc  aaa  gcg  cgc  atc  atc  cgc  tat  ttc  tac  aat  gct  aaa  gca

  46   47   48   49   50   51   52   53   54   55   56   57   58   59   60
   G    L    C    Q    T    F    V    Y    G    G    C    R    A    K    R
 ggc  ctg  tgc  cag  acc  ttt  gta  tac  ggt  ggt  tgc  cgt  gct  aag  cgt

  61   62   63   64   65   66   67   68   69   70   71   72   73   74   75
   N    N    F    K    S    A    E    D    C    M    R    T    C    G    G
 aac  aac  ttt  aaa  tcg  gcc  gaa  gat  tgc  atg  cgt  acc  tgc  ggt  ggc

  76   77   78   79   80   81   82   83   84   85   86   87   88   89   90
   A    G    A    A    E    T    V    E    S    C    L    A    K    P    H
 gcc  GGC  GCC  GCT  GAA  ACT  GTT  GAA  AGT  TGT  TTA  GCA  AAA  CCC  CAT
     |<------- mature gene III protein --------------------

  91   92   93   94   95   96   97   98   99  100  101  102  103  104  105
   T    E    N    S    F    T    N    V    W    K    D    D    K    T    L
 ACA  GAA  AAT  TCA  TTT  ACT  AAC  GTC  TGG  AAA  GAC  GAC  AAA  ACT  TTA
```

Table 116: translation of <u>Signal-III</u>::<u>bpti</u>::<u>mature-III</u>
(continued)

```
106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
 D   R   Y   A   N   Y   E   G   C   L   W   N   A   T   G
GAT CGT TAC GCT AAC TAT GAG GGT TGT CTG TGG AAT GCT ACA GGC

121 122 123 124 125 126 127 128 129 130 131 132 133 134 135
 V   V   V   C   T   G   D   E   T   Q   C   Y   G   T   W
GTT GTA GTT TGT ACT GGT GAC GAA ACT CAG TGT TAC GGT ACA TGG

136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
 V   P   I   G   L   A   I   P   E   N   E   G   G   G   S
GTT CCT ATT GGG CTT GCT ATC CCT GAA AAT GAG GGT GGT GGC TCT

151 152 153 154 155 156 157 158 159 160 161 162 163 164 165
 E   G   G   G   S   E   G   G   G   S   E   G   G   G   T
GAG GGT GGC GGT TCT GAG GGT GGC GGT TCT GAG GGT GGC GGT ACT

166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
 K   P   P   E   Y   G   D   T   P   I   P   G   Y   T   Y
AAA CCT CCT GAG TAC GGT GAT ACA CCT ATT CCG GGC TAT ACT TAT

181 182 183 184 185 186 187 188 189 190 191 192 193 194 195
 I   N   P   L   D   G   T   Y   P   P   G   T   E   Q   N
ATC AAC CCT CTC GAC GGC ACT TAT CCG CCT GGT ACT GAG CAA AAC

196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
 P   A   N   P   N   P   S   L   E   E   S   Q   P   L   N
CCC GCT AAT CCT AAT CCT TCT CTT GAG GAG TCT CAG CCT CTT AAT

211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
 T   F   M   F   Q   N   N   R   F   R   N   R   Q   G   A
ACT TTC ATG TTT CAG AAT AAT AGG TTC CGA AAT AGG CAG GGG GCA
```

126

Table 116: translation of <u>Signal-III</u>::<u>bpti</u>::<u>mature-III</u>
(continued)

```
226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
 L   T   V   Y   T   G   T   V   T   Q   G   T   D   P   V
TTA ACT GTT TAT ACG GGC ACT GTT ACT CAA GGC ACT GAC CCC GTT

241 242 243 244 245 246 247 248 249 250 251 252 253 254 255
 K   T   Y   Y   Q   Y   T   P   V   S   S   K   A   M   Y
AAA ACT TAT TAC CAG TAC ACT CCT GTA TCA TCA AAA GCC ATG TAT

256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
 D   A   Y   W   N   G   K   F   R   D   C   A   F   H   S
GAC GCT TAC TGG AAC GGT AAA TTC AGA GAC TGC GCT TTC CAT TCT

271 272 273 274 275 276 277 278 279 280 281 282 283 284 285
 G   F   N   E   D   P   F   V   C   E   Y   Q   G   Q   S
GGC TTT AAT GAG GAT CCA TTC GTT TGT GAA TAT CAA GGC CAA TCG

286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
 S   D   L   P   Q   P   P   V   N   A   G   G   G   S   G
TCT GAC CTG CCT CAA CCT CCT GTC AAT GCT GGC GGC GGC TCT GGT

301 302 303 304 305 306 307 308 309 310 311 312 313 314 315
 G   G   S   G   G   G   S   E   G   G   G   S   E   G   G
GGT GGT TCT GGT GGC GGC TCT GAG GGT GGT GGC TCT GAG GGT GGC

316 317 318 319 320 321 322 323 324 325 326 327 328 329 330
 G   S   E   G   G   G   S   E   G   G   G   S   G   G   G
GGT TCT GAG GGT GGC GGC TCT GAG GGA GGC GGT TCC GGT GGT GGC

331 332 333 334 335 336 337 338 339 340 341 342 343 344 345
 S   G   S   G   D   F   D   Y   E   K   M   A   N   A   N
TCT GGT TCC GGT GAT TTT GAT TAT GAA AAG ATG GCA AAC GCT AAT
```

Table 116: translation of Signal-III::bpti::mature-III (continued)

```
346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
 K   G   A   M   T   E   N   A   D   E   N   A   L   Q   S
AAG GGG GCT ATG ACC GAA AAT GCC GAT GAA AAC GCG CTA CAG TCT

361 362 363 364 365 366 367 368 369 370 371 372 373 374 375
 D   A   K   G   K   L   D   S   V   A   T   D   Y   G   A
GAC GCT AAA GGC AAA CTT GAT TCT GTC GCT ACT GAT TAC GGT GCT

376 377 378 379 380 381 382 383 384 385 386 387 388 389 390
 A   I   D   G   F   I   G   D   V   S   G   L   A   N   G
GCT ATC GAT GGT TTC ATT GGT GAC GTT TCC GGC CTT GCT AAT GGT

391 392 393 394 395 396 397 398 399 400 401 402 403 404 405
 N   G   A   T   G   D   F   A   G   S   N   S   Q   M   A
AAT GGT GCT ACT GGT GAT TTT GCT GGC TCT AAT TCC CAA ATG GCT

406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
 Q   V   G   D   G   D   N   S   P   L   M   N   N   F   R
CAA GTC GGT GAC GGT GAT AAT TCA CCT TTA ATG AAT AAT TTC CGT

421 422 423 424 425 426 427 428 429 430 431 432 433 434 435
 Q   Y   L   P   S   L   P   Q   S   V   E   C   R   P   F
CAA TAT TTA CCT TCC CTC CCT CAA TCG GTT GAA TGT CGC CCT TTT

436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
 V   F   S   A   G   K   P   Y   E   F   S   I   D   C   D
GTC TTT AGC GCT GGT AAA CCA TAT GAA TTT TCT ATT GAT TGT GAC
```

128

Table 116: translation of <u>Signal-III</u>::<u>bpti</u>::<u>mature-III</u>
(continued)

```
451 452 453 454 455 456 457 458 459 460 461 462 463 464 465
 K   I   N   L   F   R   G   V   F   A   F   L   L   Y   V
AAA ATA AAC TTA TTC CGT GGT GTC TTT GCG TTT CTT TTA TAT GTT
                        |<----- uncharged anchor region -----

466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
 A   T   F   M   Y   V   F   S   T   F   A   N   I   L   R
GCC ACC TTT ATG TAT GTA TTT TCT ACG TTT GCT AAC ATA CTG CGT
--------- uncharged anchor region continues --------->|

481 482 483 484 485
 N   K   E   S   .
AAT AAG GAG TCT TAA
```

```
Molecular weight of peptide  =      58884
Charge on peptide            =        -20
[A+G+P]                      =        143
[C+F+H+I+L+M+V+W+Y]          =        140
[D+E+K+R+N+Q+S+T+.]          =        202
```

129

Table 116: translation of <u>Signal-III</u>::<u>bpti</u>::<u>mature-III</u>
(continued)

	Second Base				
	t	c	a	g	
t	15	21	15	8	t
	12	5	10	6	c
	10	4	0	0	a
	0	3	0	4	g
c	6	20	2	8	t
	3	4	0	3	c
	1	4	9	1	a
	4	3	7	0	g
a	5	19	21	1	t
	5	4	11	1	c
	2	4	16	1	a
	8	2	4	2	g
g	13	22	14	41	t
	6	7	12	29	c
	4	5	12	1	a
	1	3	16	4	g

AA	#	AA	#	AA	#	AA	#
A	37	C	14	D	26	E	28
F	27	G	75	H	2	I	12
K	20	L	24	M	9	N	32
P	31	Q	16	R	15	S	35
T	29	V	23	W	4	Y	25
.	1						

130

Table 130: Sampling of a Library encoded by $(NNK)^6$

A.   Numbers of hexapeptides in each class

5   total   =   64,000,000 stop-free sequences.

α can be one of [WMFYCIKDENHQ]
Φ can be one of [PTAVG]
Ω can be one of [SLR]

10

$\alpha\alpha\alpha\alpha\alpha\alpha$	=	2985984.	$\Phi\alpha\alpha\alpha\alpha\alpha$	=	7464960.
$\Omega\alpha\alpha\alpha\alpha\alpha$	=	4478976.	$\Phi\Phi\alpha\alpha\alpha\alpha$	=	7776000.
$\Phi\Omega\alpha\alpha\alpha\alpha$	=	9331200.	$\Omega\Omega\alpha\alpha\alpha\alpha$	=	2799360.
$\Phi\Phi\Phi\alpha\alpha\alpha$	=	4320000.	$\Phi\Phi\Omega\alpha\alpha\alpha$	=	7776000.
$\Phi\Omega\Omega\alpha\alpha\alpha$	=	4665600.	$\Omega\Omega\Omega\alpha\alpha\alpha$	=	933120.
$\Phi\Phi\Phi\Phi\alpha\alpha$	=	1350000.	$\Phi\Phi\Phi\Omega\alpha\alpha$	=	3240000.
$\Phi\Phi\Omega\Omega\alpha\alpha$	=	2916000.	$\Phi\Omega\Omega\Omega\alpha\alpha$	=	1166400.
$\Omega\Omega\Omega\Omega\alpha\alpha$	=	174960.	$\Phi\Phi\Phi\Phi\Phi\alpha$	=	225000.
$\Phi\Phi\Phi\Phi\Omega\alpha$	=	675000.	$\Phi\Phi\Phi\Omega\Omega\alpha$	=	810000.
$\Phi\Phi\Omega\Omega\Omega\alpha$	=	486000.	$\Phi\Omega\Omega\Omega\Omega\alpha$	=	145800.
$\Omega\Omega\Omega\Omega\Omega\alpha$	=	17496.	$\Phi\Phi\Phi\Phi\Phi\Phi$	=	15625.
$\Phi\Phi\Phi\Phi\Phi\Omega$	=	56250.	$\Phi\Phi\Phi\Phi\Omega\Omega$	=	84375.
$\Phi\Phi\Phi\Omega\Omega\Omega$	=	67500.	$\Phi\Phi\Omega\Omega\Omega\Omega$	=	30375.
$\Phi\Omega\Omega\Omega\Omega\Omega$	=	7290.	$\Omega\Omega\Omega\Omega\Omega\Omega$	=	729.

25

$\Phi\Phi\Omega\Omega\alpha\alpha$, for example, stands for the set of peptides having
two amino acids from the α class, two from Φ, and two from
Ω arranged in any order. There are, for example, $729 = 3^6$
sequences composed entirely of S, L, and R.

30

Table 130: Sampling of a Library encoded by $(NNK)^6$
(continued)

B.   Probability that any given stop-free DNA sequence will
encode a hexapeptide from a stated class.

	P	% of class
$\alpha\alpha\alpha\alpha\alpha\alpha$...	3.364E-03	(1.13E-07)
$\Phi\alpha\alpha\alpha\alpha\alpha$...	1.682E-02	(2.25E-07)
$\Omega\alpha\alpha\alpha\alpha\alpha$...	1.514E-02	(3.38E-07)
$\Phi\Phi\alpha\alpha\alpha\alpha$...	3.505E-02	(4.51E-07)
$\Phi\Omega\alpha\alpha\alpha\alpha$...	6.308E-02	(6.76E-07)
$\Omega\Omega\alpha\alpha\alpha\alpha$...	2.839E-02	(1.01E-06)
$\Phi\Phi\Phi\alpha\alpha\alpha$...	3.894E-02	(9.01E-07)
$\Phi\Phi\Omega\alpha\alpha\alpha$...	1.051E-01	(1.35E-06)
$\Phi\Omega\Omega\alpha\alpha\alpha$...	9.463E-02	(2.03E-06)
$\Omega\Omega\Omega\alpha\alpha\alpha$...	2.839E-02	(3.04E-06)
$\Phi\Phi\Phi\Phi\alpha\alpha$...	2.434E-02	(1.80E-06)
$\Phi\Phi\Phi\Omega\alpha\alpha$...	8.762E-02	(2.70E-06)
$\Phi\Phi\Omega\Omega\alpha\alpha$...	1.183E-01	(4.06E-06)
$\Phi\Omega\Omega\Omega\alpha\alpha$...	7.097E-02	(6.08E-06)
$\Omega\Omega\Omega\Omega\alpha\alpha$...	1.597E-02	(9.13E-06)
$\Phi\Phi\Phi\Phi\Phi\alpha$...	8.113E-03	(3.61E-06)
$\Phi\Phi\Phi\Phi\Omega\alpha$...	3.651E-02	(5.41E-06)
$\Phi\Phi\Phi\Omega\Omega\alpha$...	6.571E-02	(8.11E-06)
$\Phi\Phi\Omega\Omega\Omega\alpha$...	5.914E-02	(1.22E-05)
$\Phi\Omega\Omega\Omega\Omega\alpha$...	2.661E-02	(1.83E-05)
$\Omega\Omega\Omega\Omega\Omega\alpha$...	4.790E-03	(2.74E-05)
$\Phi\Phi\Phi\Phi\Phi\Phi$...	1.127E-03	(7.21E-06)
$\Phi\Phi\Phi\Phi\Phi\Omega$...	6.084E-03	(1.08E-05)
$\Phi\Phi\Phi\Phi\Omega\Omega$...	1.369E-02	(1.62E-05)
$\Phi\Phi\Phi\Omega\Omega\Omega$...	1.643E-02	(2.43E-05)
$\Phi\Phi\Omega\Omega\Omega\Omega$...	1.109E-02	(3.65E-05)
$\Phi\Omega\Omega\Omega\Omega\Omega$...	3.992E-03	(5.48E-05)
$\Omega\Omega\Omega\Omega\Omega\Omega$...	5.988E-04	(8.21E-05)

132

Table 130: Sampling of a Library encoded by (NNK)[6]
(continued)

C.   Number of different stop-free amino-acid sequences in
each class expected for various library sizes

Library size =   1.0000E+06

total =   9.7446E+05   % sampled =   1.52

Class	Number	%	Class	Number	%
$\alpha\alpha\alpha\alpha\alpha\alpha$...	3362.6(	.1)	$\Phi\alpha\alpha\alpha\alpha\alpha$...	16803.4(	.2)
$\Omega\alpha\alpha\alpha\alpha\alpha$...	15114.6(	.3)	$\Phi\Phi\alpha\alpha\alpha\alpha$...	34967.8(	.4)
$\Phi\Omega\alpha\alpha\alpha\alpha$...	62871.1(	.7)	$\Omega\Omega\alpha\alpha\alpha\alpha$...	28244.3(	1.0)
$\Phi\Phi\Phi\alpha\alpha\alpha$...	38765.7(	.9)	$\Phi\Phi\Omega\alpha\alpha\alpha$...	104432.2(	1.3)
$\Phi\Omega\Omega\alpha\alpha\alpha$...	93672.7(	2.0)	$\Omega\Omega\Omega\alpha\alpha\alpha$...	27960.3(	3.0)
$\Phi\Phi\Phi\Phi\alpha\alpha$...	24119.9(	1.8)	$\Phi\Phi\Phi\Omega\alpha\alpha$...	86442.5(	2.7)
$\Phi\Phi\Omega\Omega\alpha\alpha$...	115915.5(	4.0)	$\Phi\Omega\Omega\Omega\alpha\alpha$...	68853.5(	5.9)
$\Omega\Omega\Omega\Omega\alpha\alpha$...	15261.1(	8.7)	$\Phi\Phi\Phi\Phi\Phi\alpha$...	7968.1(	3.5)
$\Phi\Phi\Phi\Phi\Omega\alpha$...	35537.2(	5.3)	$\Phi\Phi\Phi\Omega\Omega\alpha$...	63117.5(	7.8)
$\Phi\Phi\Omega\Omega\Omega\alpha$...	55684.4(	11.5)	$\Phi\Omega\Omega\Omega\Omega\alpha$...	24325.9(	16.7)
$\Omega\Omega\Omega\Omega\Omega\alpha$...	4190.6(	24.0)	$\Phi\Phi\Phi\Phi\Phi\Phi$...	1087.1(	7.0)
$\Phi\Phi\Phi\Phi\Phi\Omega$...	5767.0(	10.3)	$\Phi\Phi\Phi\Phi\Omega\Omega$...	12637.2(	15.0)
$\Phi\Phi\Phi\Omega\Omega\Omega$...	14581.7(	21.6)	$\Phi\Phi\Omega\Omega\Omega\Omega$...	9290.2(	30.6)
$\Phi\Omega\Omega\Omega\Omega\Omega$...	3073.9(	42.2)	$\Omega\Omega\Omega\Omega\Omega\Omega$...	408.4(	56.0)

Library size =   3.0000E+06

total =   2.7885E+06   % sampled =   4.36

Class	Number	%	Class	Number	%
$\alpha\alpha\alpha\alpha\alpha\alpha$...	10076.4(	.3)	$\Phi\alpha\alpha\alpha\alpha\alpha$...	50296.9(	.7)
$\Omega\alpha\alpha\alpha\alpha\alpha$...	45190.9(	1.0)	$\Phi\Phi\alpha\alpha\alpha\alpha$...	104432.2(	1.3)
$\Phi\Omega\alpha\alpha\alpha\alpha$...	187345.5(	2.0)	$\Omega\Omega\alpha\alpha\alpha\alpha$...	83880.9(	3.0)
$\Phi\Phi\Phi\alpha\alpha\alpha$...	115256.6(	2.7)	$\Phi\Phi\Omega\alpha\alpha\alpha$...	309107.9(	4.0)
$\Phi\Omega\Omega\alpha\alpha\alpha$...	275413.9(	5.9)	$\Omega\Omega\Omega\alpha\alpha\alpha$...	81392.5(	8.7)
$\Phi\Phi\Phi\Phi\alpha\alpha$...	71074.5(	5.3)	$\Phi\Phi\Phi\Omega\alpha\alpha$...	252470.2(	7.8)
$\Phi\Phi\Omega\Omega\alpha\alpha$...	334106.2(	11.5)	$\Phi\Omega\Omega\Omega\alpha\alpha$...	194606.9(	16.7)
$\Omega\Omega\Omega\Omega\alpha\alpha$...	41905.9(	24.0)	$\Phi\Phi\Phi\Phi\Phi\alpha$...	23067.8(	10.3)
$\Phi\Phi\Phi\Phi\Omega\alpha$...	101097.3(	15.0)	$\Phi\Phi\Phi\Omega\Omega\alpha$...	174981.0(	21.6)
$\Phi\Phi\Omega\Omega\Omega\alpha$...	148643.7(	30.6)	$\Phi\Omega\Omega\Omega\Omega\alpha$...	61478.9(	42.2)
$\Omega\Omega\Omega\Omega\Omega\alpha$...	9801.0(	56.0)	$\Phi\Phi\Phi\Phi\Phi\Phi$...	3039.6(	19.5)
$\Phi\Phi\Phi\Phi\Phi\Omega$...	15587.7(	27.7)	$\Phi\Phi\Phi\Phi\Omega\Omega$...	32516.8(	38.5)
$\Phi\Phi\Phi\Omega\Omega\Omega$...	34975.6(	51.8)	$\Phi\Phi\Omega\Omega\Omega\Omega$...	20215.5(	66.6)
$\Phi\Omega\Omega\Omega\Omega\Omega$...	5879.9(	80.7)	$\Omega\Omega\Omega\Omega\Omega\Omega$...	667.0(	91.5)

Table 130: Sampling of a Library encoded by (NNK)^6
(continued)

Library size =   1.0000E+07

total =   8.1204E+06 % sampled =   12.69

ααααα...	33455.9(| 1.1)	Φαααα...	166342.4(| 2.2)
Ωαααα...	148871.1(| 3.3)	ΦΦαααα...	342685.7(| 4.4)
ΦΩαααα...	609987.6(| 6.5)	ΩΩαααα...	269958.3(| 9.6)
ΦΦΦαααα...	372371.8(| 8.6)	ΦΦΩαααα...	983416.4(| 12.6)
ΦΩΩααα...	856471.6(| 18.4)	ΩΩΩααα...	244761.5(| 26.2)
ΦΦΦΦαα...	222702.0(| 16.5)	ΦΦΦΩαα...	767692.5(| 23.7)
ΦΦΩΩαα...	972324.6(| 33.3)	ΦΩΩΩαα...	531651.3(| 45.6)
ΩΩΩΩαα...	104722.3(| 59.9)	ΦΦΦΦΦα...	68111.0(| 30.3)
ΦΦΦΦΩα...	281976.3(| 41.8)	ΦΦΦΩΩα...	450120.2(| 55.6)
ΦΦΩΩΩα...	342072.1(| 70.4)	ΦΩΩΩΩα...	122302.6(| 83.9)
ΩΩΩΩΩα...	16364.0(| 93.5)	ΦΦΦΦΦΦ...	8028.0(| 51.4)
ΦΦΦΦΦΩ...	37179.9(| 66.1)	ΦΦΦΦΩΩ...	67719.5(| 80.3)
ΦΦΦΩΩΩ...	61580.0(| 91.2)	ΦΦΩΩΩΩ...	29586.1(| 97.4)
ΦΩΩΩΩΩ...	7259.5(| 99.6)	ΩΩΩΩΩΩ...	728.8(100.0)

Library size =   3.0000E+07

total =   1.8633E+07 % sampled =   29.11

ααααα...	99247.4(| 3.3)	Φαααα...	487990.0(| 6.5)
Ωαααα...	431933.3(| 9.6)	ΦΦαααα...	983416.5(| 12.6)
ΦΩαααα...	1712943.0(| 18.4)	ΩΩαααα...	734284.6(| 26.2)
ΦΦΦαααα...	1023590.0(| 23.7)	ΦΦΩαααα...	2592866.0(| 33.3)
ΦΩΩααα...	2126605.0(| 45.6)	ΩΩΩααα...	558519.0(| 59.9)
ΦΦΦΦαα...	563952.6(| 41.8)	ΦΦΦΩαα...	1800481.0(| 55.6)
ΦΦΩΩαα...	2052433.0(| 70.4)	ΦΩΩΩαα...	978420.5(| 83.9)
ΩΩΩΩαα...	163640.3(| 93.5)	ΦΦΦΦΦα...	148719.7(| 66.1)
ΦΦΦΦΩα...	541755.7(| 80.3)	ΦΦΦΩΩα...	738960.1(| 91.2)
ΦΦΩΩΩα...	473377.0(| 97.4)	ΦΩΩΩΩα...	145189.7(| 99.6)
ΩΩΩΩΩα...	17491.3(100.0)	ΦΦΦΦΦΦ...	13829.1(| 88.5)
ΦΦΦΦΦΩ...	54058.1(| 96.1)	ΦΦΦΦΩΩ...	83726.0(| 99.2)
ΦΦΦΩΩΩ...	67454.5(| 99.9)	ΦΦΩΩΩΩ...	30374.5(100.0)
ΦΩΩΩΩΩ...	7290.0(100.0)	ΩΩΩΩΩΩ...	729.0(100.0)

134

Table 130: Sampling of a Library encoded by (NNK)[6]
(continued)

Library size =   7.6000E+07

   total =   3.2125E+07 % sampled =   50.19

```
αααααα...   245057.8(  8.2)   Φαααα...  1175010.0( 15.7)
Ωααααα...  1014733.0( 22.7)   ΦΦαααα...  2255280.0( 29.0)
ΦΩαααα...  3749112.0( 40.2)   ΩΩαααα...  1504128.0( 53.7)
ΦΦΦααα...  2142478.0( 49.6)   ΦΦΩααα...  4993247.0( 64.2)
ΦΩΩααα...  3666785.0( 78.6)   ΩΩΩααα...   840691.9( 90.1)
ΦΦΦΦαα...  1007002.0( 74.6)   ΦΦΦΩαα...  2825063.0( 87.2)
ΦΦΩΩαα...  2782358.0( 95.4)   ΦΩΩΩαα...  1154956.0( 99.0)
ΩΩΩΩαα...   174790.0( 99.9)   ΦΦΦΦΦα...   210475.6( 93.5)
ΦΦΦΦΩα...   663929.3( 98.4)   ΦΦΦΩΩα...   808298.6( 99.8)
ΦΩΩΩΩα...   485953.2(100.0)   ΦΩΩΩΩα...   145799.9(100.0)
ΩΩΩΩΩα...    17496.0(100.0)   ΦΦΦΦΦΦ...    15559.9( 99.6)
ΦΦΦΦΦΩ...    56234.9(100.0)   ΦΦΦΦΩΩ...    84374.6(100.0)
ΦΦΦΩΩΩ...    67500.0(100.0)   ΦΩΩΩΩΩ...    30375.0(100.0)
ΦΩΩΩΩΩ...     7290.0(100.0)   ΩΩΩΩΩΩ...      729.0(100.0)
```

Library size =   1.0000E+08

   total =   3.6537E+07 % sampled =   57.09

```
αααααα...   318185.1( 10.7)   Φαααα...  1506161.0( 20.2)
Ωααααα...  1284677.0( 28.7)   ΦΦαααα...  2821285.0( 36.3)
ΦΩαααα...  4585163.0( 49.1)   ΩΩαααα...  1783932.0( 63.7)
ΦΦΦααα...  2566085.0( 59.4)   ΦΦΩααα...  5764391.0( 74.1)
ΦΩΩααα...  4051713.0( 86.8)   ΩΩΩααα...   888584.3( 95.2)
ΦΦΦΦαα...  1127473.0( 83.5)   ΦΦΦΩαα...  3023170.0( 93.3)
ΦΦΩΩαα...  2865517.0( 98.3)   ΦΩΩΩαα...  1163743.0( 99.8)
ΩΩΩΩαα...   174941.0(100.0)   ΦΦΦΦΦα...   218886.6( 97.3)
ΦΦΦΦΩα...   671976.9( 99.6)   ΦΦΦΩΩα...   809757.3(100.0)
ΦΩΩΩΩα...   485997.5(100.0)   ΦΩΩΩΩα...   145800.0(100.0)
ΩΩΩΩΩα...    17496.0(100.0)   ΦΦΦΦΦΦ...    15613.5( 99.9)
ΦΦΦΦΦΩ...    56248.9(100.0)   ΦΦΦΦΩΩ...    84375.0(100.0)
ΦΦΦΩΩΩ...    67500.0(100.0)   ΦΩΩΩΩΩ...    30375.0(100.0)
ΦΩΩΩΩΩ...     7290.0(100.0)   ΩΩΩΩΩΩ...      729.0(100.0)
```

Table 130: Sampling of a Library encoded by (NNK)[6]
(continued)

Library size =   3.0000E+08

5

total =   5.2634E+07 % sampled =   82.24

ααααα...   856451.3(28.7)   Φαααα... 3668130.0(49.1)
Ωαααα... 2854291.0(63.7)   ΦΦαααα... 5764391.0(74.1)
10   ΦΩαααα... 8103426.0(86.8)   ΩΩαααα... 2665753.0(95.2)
ΦΦΦαα... 4030893.0(93.3)   ΦΦΩαα... 7641378.0(98.3)
ΦΩΩαα... 4654972.0(99.8)   ΩΩΩαα... 933018.6(100.0)
ΦΦΦΦα... 1343954.0(99.6)   ΦΦΦΩαα... 3239029.0(100.0)
ΦΦΩΩαα... 2915985.0(100.0)   ΦΩΩΩαα... 1166400.0(100.0)
15   ΩΩΩΩαα...   174960.0(100.0)   ΦΦΦΦα... 224995.5(100.0)
ΦΦΦΦΩα...   674999.9(100.0)   ΦΦΦΩΩα... 810000.0(100.0)
ΦΦΩΩα...   486000.0(100.0)   ΦΩΩΩΩα... 145800.0(100.0)
ΩΩΩΩα...   17496.0(100.0)   ΦΦΦΦΦ...   15625.0(100.0)
ΦΦΦΦΦΩ...   56250.0(100.0)   ΦΦΦΦΩΩ...   84375.0(100.0)
20   ΦΦΦΩΩΩ...   67500.0(100.0)   ΦΦΩΩΩΩ...   30375.0(100.0)
ΦΩΩΩΩΩ...   7290.0(100.0)   ΩΩΩΩΩΩ...   729.0(100.0)

Library size =   1.0000E+09

25

total =   6.1999E+07 % sampled =   96.87

ααααα... 2018278.0(67.6)   Φαααα... 6680917.0(89.5)
Ωαααα... 4326519.0(96.6)   ΦΦαααα... 7690221.0(98.9)
30   ΦΩαααα... 9320389.0(99.9)   ΩΩαααα... 2799250.0(100.0)
ΦΦΦαα... 4319475.0(100.0)   ΦΦΩαα... 7775990.0(100.0)
ΦΩΩαα... 4665600.0(100.0)   ΩΩΩαα... 933120.0(100.0)
ΦΦΦΦα... 1350000.0(100.0)   ΦΦΦΩαα... 3240000.0(100.0)
ΦΦΩΩαα... 2916000.0(100.0)   ΦΩΩΩαα... 1166400.0(100.0)
35   ΩΩΩΩαα...   174960.0(100.0)   ΦΦΦΦα... 225000.0(100.0)
ΦΦΦΦΩα...   675000.0(100.0)   ΦΦΦΩΩα... 810000.0(100.0)
ΦΦΩΩΩα...   486000.0(100.0)   ΦΩΩΩΩα... 145800.0(100.0)
ΩΩΩΩΩα...   17496.0(100.0)   ΦΦΦΦΦ...   15625.0(100.0)
ΦΦΦΦΩ...   56250.0(100.0)   ΦΦΦΦΩΩ...   84375.0(100.0)
40   ΦΦΦΩΩΩ...   67500.0(100.0)   ΦΦΩΩΩΩ...   30375.0(100.0)
ΦΩΩΩΩΩ...   7290.0(100.0)   ΩΩΩΩΩΩ...   729.0(100.0)

136

Table 130: Sampling of a Library encoded by (NNK)^6
(continued)

Library size =   3.0000E+09

5   total =   6.3890E+07   % sampled =   99.83

ααααα...	2884346.0 (96.6)	Φαααα... 7456311.0 (99.9)
Ωαααα...	4478800.0 (100.0)	ΦΦαααα... 7775990.0 (100.0)
ΦΩαααα...	9331200.0 (100.0)	ΩΩαααα... 2799360.0 (100.0)
ΦΦΦααα...	4320000.0 (100.0)	ΦΦΩααα... 7776000.0 (100.0)
ΦΩΩααα...	4665600.0 (100.0)	ΩΩΩααα... 933120.0 (100.0)
ΦΦΦΦαα...	1350000.0 (100.0)	ΦΦΦΩαα... 3240000.0 (100.0)
ΦΦΩΩαα...	2916000.0 (100.0)	ΦΩΩΩαα... 1166400.0 (100.0)
ΩΩΩΩαα...	174960.0 (100.0)	ΦΦΦΦΦα... 225000.0 (100.0)
ΦΦΦΦΩα...	675000.0 (100.0)	ΦΦΦΩΩα... 810000.0 (100.0)
ΦΦΩΩΩα...	486000.0 (100.0)	ΦΩΩΩΩα... 145800.0 (100.0)
ΩΩΩΩΩα...	17496.0 (100.0)	ΦΦΦΦΦΦ... 15625.0 (100.0)
ΦΦΦΦΦΩ...	56250.0 (100.0)	ΦΦΦΦΩΩ... 84375.0 (100.0)
ΦΦΦΩΩΩ...	67500.0 (100.0)	ΦΦΩΩΩΩ... 30375.0 (100.0)
ΦΩΩΩΩΩ...	7290.0 (100.0)	ΩΩΩΩΩΩ... 729.0 (100.0)

Table 130, continued

D. Formulae for tabulated quantities.

5   Lsize is the number of independent transformants.
    31**6 is 31 to sixth power; 6*3 means 6 times 3.
    A = Lsize/(31**6)
    α can be one of [WMFYCIKDENHQ.]
    Φ can be one of [PTAVG]
10 Ω can be one of [SLR]
    F0 = (12)**6   F1 = (12)**5   F2 = (12)**4
    F3 = (12)**3   F4 = (12)**2   F5 = (12)
    F6 = 1

15 ααααα = F0 * (1-exp(-A))
    Φααααα = 6 * 5 * F1 * (1-exp(-2*A))
    Ωααααα = 6 * 3 * F1 * (1-exp(-3*A))
    ΦΦαααα = (15) * 5**2 * F2 * (1-exp(-4*A))
    ΦΩαααα = (6*5)*5*3 *F2 * (1-exp(-6*A))
20 ΩΩαααα = (15) * 3**2 * F2 * (1-exp(-9*A))
    ΦΦΦααα = (20)*(5**3) * F3 * (1-exp(-8*A))
    ΦΦΩααα = (60)*(5*5*3)*F3* (1-exp(-12*A))
    ΦΩΩααα = (60)*(5*3*3)*F3*(1-exp(-18*A))
    ΩΩΩααα = (20)*(3)**3*F3*(1-exp(-27*A))
25 ΦΦΦΦαα = (15)*(5)**4*F4*(1-exp(-16*A))
    ΦΦΦΩαα = (60)*(5)**3*3*F4*(1-exp(-24*A))
    ΦΦΩΩαα = (90)*(5*5*3*3)*F4*(1-exp(-36*A))
    ΦΩΩΩαα = (60)*(5*3*3*3)*F4*(1-exp(-54*A))
    ΩΩΩΩαα = (15)*(3)**4 * F4 *(1-exp(-81*A))
30 ΦΦΦΦΦα = (6)*(5)**5 * F5 * (1-exp(-32*A))
    ΦΦΦΦΩα = 30*5*5*5*3*F5*(1-exp(-48*A))
    ΦΦΦΩΩα = 60*5*5*5*3*3*F5*(1-exp(-72*A))
    ΦΦΩΩΩα = 60*5*5*3*3*3*F5*(1-exp(-108*A))
    ΦΩΩΩΩα = 30*5*3*3*3*3*F5*(1-exp(-162*A))
35 ΩΩΩΩΩα = 6*3*3*3*3*3*F5*(1-exp(-243*A))
    ΦΦΦΦΦΦ = 5**6 * (1-exp(-64*A))
    ΦΦΦΦΦΩ = 6*3*5**5*(1-exp(-96*A))
    ΦΦΦΦΩΩ = 15*3*3*5**4*(1-exp(-144*A))
    ΦΦΦΩΩΩ = 20*3**3*5**3*(1-exp(-216*A))
40 ΦΦΩΩΩΩ = 15*3**4*5**2*(1-exp(-324*A))
    ΦΩΩΩΩΩ = 6*3**5*5*(1-exp(-486*A))
    ΩΩΩΩΩΩ = 3**6*(1-exp(-729*A))
    total = ααααα + Φααααα + Ωααααα + ΦΦαααα + ΦΩαααα +
            ΩΩαααα + ΦΦΦααα + ΦΦΩααα + ΦΩΩααα + ΩΩΩααα +
45   ΦΦΦΦαα + ΦΦΦΩαα + ΦΦΩΩαα + ΦΩΩΩαα + ΩΩΩΩαα +
            ΦΦΦΦΦα + ΦΦΦΦΩα + ΦΦΦΩΩα + ΦΦΩΩΩα + ΦΩΩΩΩα +
            ΩΩΩΩΩα + ΦΦΦΦΦΦ + ΦΦΦΦΦΩ + ΦΦΦΦΩΩ + ΦΦΦΩΩΩ +
            ΦΦΩΩΩΩ + ΦΩΩΩΩΩ + ΩΩΩΩΩΩ

138

Table 131: Sampling of a Library
Encoded by $(NNT)^4(NNG)^2$

X can be F,S,Y,C,L,P,H,R,I,T,N,V,A,D,G

Γ can be L^2,R^2,S,W,P,Q,M,T,K,V,A,E,G

Library comprises $8.55 \cdot 10^6$ amino-acid sequences; $1.47 \cdot 10^7$ DNA sequences.

Total number of possible aa sequences=   8,555,625

x	LVPTARGFYCHIND
S	S
θ	VPTAGWQMKES
Ω	LR

The first, second, fifth, and sixth positions can hold x or S; the third and fourth position can hold θ or Ω. I have lumped sequences by the number of xs, Ss, θs, and Ωs.

For example xxθΩSS stands for:
[xxθΩSS, xSθΩxS, xSθΩSx, SSθΩxx, SxθΩxS, SxθΩSx,
 xxΩθSS, xSΩθxS, xSΩθSx, SSΩθxx, SxΩθxS, SxΩθSx]

The following table shows the likelihood that any particular DNA sequence will fall into one of the defined classes.

Library size =   1.0   Sampling = .00001%

total........	1.0000E+00	%sampled.....	1.1688E-07
xxθθxx.......	3.1524E-01	xxθΩxx.......	2.2926E-01
xxΩΩxx.......	4.1684E-02	xxθθxS.......	1.8013E-01
xxθΩxS.......	1.3101E-01	xxΩΩxS.......	2.3819E-02
xxθθSS.......	3.8600E-02	xxθΩSS.......	2.8073E-02
xxΩΩSS.......	5.1042E-03	xSθθSS.......	3.6762E-03
xSθΩSS.......	2.6736E-03	xSΩΩSS.......	4.8611E-04
SSθθSS.......	1.3129E-04	SSθΩSS.......	9.5486E-05
SSΩΩSS.......	1.7361E-05		

139

Table 131: Sampling of a Library
Encoded by $(NNT)^4(NNG)^2$
(continued)

5　　　The following sections show how many sequences of each class are expected for libraries of different sizes.

10　　Library size ＝　　　1.0000E+05

total......	9.9137E+04		fraction sampled ＝ 1.1587E-02		
Type	Number	%	Type	Number	%
xxθθxx.....	31416.9(	.7)	xxθΩxx.....	22771.4(	1.3)
xxΩΩxx.....	4112.4(	2.7)	xxθθxS.....	17891.8(	1.3)
xxθΩxS.....	12924.6(	2.7)	xxΩΩxS.....	2318.5(	5.3)
xxθθSS.....	3808.1(	2.7)	xxθΩSS.....	2732.5(	5.3)
xxΩΩSS.....	483.7(	10.3)	xSθθSS.....	357.8(	5.3)
xSθΩSS.....	253.4(	10.3)	xSΩΩSS.....	43.7(	19.5)
SSθθSS.....	12.4(	10.3)	SSθΩSS.....	8.6(	19.5)
SSΩΩSS.....	1.4(	35.2)			

Library size ＝　　　1.0000E+06

total......	9.2064E+05		fraction sampled ＝ 1.0761E-01		
xxθθxx.....	304783.9(	6.6)	xxθΩxx.....	214394.0(	12.7)
xxΩΩxx.....	36508.6(	23.8)	xxθθxS.....	168452.5(	12.7)
xxθΩxS.....	114741.4(	23.8)	xxΩΩxS.....	18383.8(	41.9)
xxθθSS.....	33807.7(	23.8)	xxθΩSS.....	21666.6(	41.9)
xxΩΩSS.....	3114.6(	66.2)	xSθθSS.....	2837.3(	41.9)
xSθΩSS.....	1631.5(	66.2)	xSΩΩSS.....	198.4(	88.6)
SSθθSS.....	80.1(	66.2)	SSθΩSS.....	39.0(	88.6)
SSΩΩSS.....	3.9(	98.7)			

Library size ＝　　3.0000E+06

total......	2.3880E+06		fraction sampled ＝ 2.7912E-01		
xxθθxx.....	855709.5(	18.4)	xxθΩxx.....	565051.6(	33.4)
xxΩΩxx.....	85564.7(	55.7)	xxθθxS.....	443969.1(	33.4)
xxθΩxS.....	268917.8(	55.7)	xxΩΩxS.....	35281.3(	80.4)
xxθθSS.....	79234.7(	55.7)	xxθΩSS.....	41581.5(	80.4)
xxΩΩSS.....	4522.6(	96.1)	xSθθSS.....	5445.2(	80.4)
xSθΩSS.....	2369.0(	96.1)	xSΩΩSS.....	223.7(	99.9)
SSθθSS.....	116.3(	96.1)	SSθΩSS.....	43.9(	99.9)
SSΩΩSS.....	4.0(100.0)				

Table 131: Sampling of a Library
Encoded by (NNT)^4(NNG)^2
(continued)

```
  5    Library size =     8.5556E+06

       total......    4.9303E+06    fraction sampled = 5.7626E-01
       xxθθxx.....  2046301.0( 44.0)  xxθΩxx..... 1160645.0( 68.7)
       xxΩΩxx.....   138575.9( 90.2)  xxθθxS.....  911935.6( 68.7)
 10    xxθΩxS.....   435524.3( 90.2)  xxΩΩxS.....   43480.7( 99.0)
       xxθθSS.....   128324.1( 90.2)  xxθΩSS.....   51245.1( 99.0)
       xxΩΩSS.....     4703.6(100.0)  xSθθSS.....    6710.7( 99.0)
       xSθΩSS.....     2463.8(100.0)  xSΩΩSS.....     224.0(100.0)
       SSθθSS.....      121.0(100.0)  SSθΩSS.....      44.0(100.0)
 15    SSΩΩSS.....        4.0(100.0)


       Library size =     1.0000E+07

       total......    5.3667E+06    fraction sampled = 6.2727E-01
 20    xxθθxx.....  2289093.0( 49.2)  xxθΩxx..... 1254877.0( 74.2)
       xxΩΩxx.....   143467.0( 93.4)  xxθθxS.....  985974.9( 74.2)
       xxθΩxS.....   450896.3( 93.4)  xxΩΩxS.....   43710.7( 99.6)
       xxθθSS.....   132853.4( 93.4)  xxθΩSS.....   51516.1( 99.6)
       xxΩΩSS.....     4703.9(100.0)  xSθθSS.....    6746.2( 99.6)
 25    xSθΩSS.....     2464.0(100.0)  xSΩΩSS.....     224.0(100.0)
       SSθθSS.....      121.0(100.0)  SSθΩSS.....      44.0(100.0)
       SSΩΩSS.....        4.0(100.0)


       Library size =     3.0000E+07
 30
       total......    7.8961E+06    fraction sampled = 9.2291E-01
       xxθθxx.....  4040589.0( 86.9)  xxθΩxx..... 1661409.0( 98.3)
       xxΩΩxx.....   153619.1(100.0)  xxθθxS..... 1305393.0( 98.3)
       xxθΩxS.....   482802.9(100.0)  xxΩΩxS.....   43904.0(100.0)
 35    xxθθSS.....   142254.4(100.0)  xxθΩSS.....   51744.0(100.0)
       xxΩΩSS.....     4704.0(100.0)  xSθθSS.....    6776.0(100.0)
       xSθΩSS.....     2464.0(100.0)  xSΩΩSS.....     224.0(100.0)
       SSθθSS.....      121.0(100.0)  SSθΩSS.....      44.0(100.0)
       SSΩΩSS.....        4.0(100.0)
```

141

Table 131: Sampling of a Library
Encoded by $(NNT)^4(NNG)^2$
(continued)

5   Library size =   5.0000E+07

       total......   8.3956E+06   fraction sampled = 9.8130E-01
       xxθθxx..... 4491779.0(96.6) xxθΩxx..... 1688387.0(99.9)
       xxΩΩxx.....   153663.8(100.0) xxθθxS..... 1326590.0(99.9)
10   xxθΩxS.....   482943.4(100.0) xxΩΩxS.....   43904.0(100.0)
       xxθθSS.....   142295.8(100.0) xxθΩSS.....   51744.0(100.0)
       xxΩΩSS.....   4704.0(100.0) xSθθSS.....   6776.0(100.0)
       xSθΩSS.....   2464.0(100.0) xSΩΩSS.....   224.0(100.0)
       SSθθSS.....   121.0(100.0) SSθΩSS.....   44.0(100.0)
15   SSΩΩSS.....   ·4.0(100.0)

       Library size =   1.0000E+08

       total......   8.5503E+06   fraction sampled = 9.9938E-01
20   xxθθxx..... 4643063.0(99.9) xxθΩxx..... 1690302.0(100.0)
       xxΩΩxx.....   153664.0(100.0) xxθθxS..... 1328094.0(100.0)
       xxθΩxS.....   482944.0(100.0) xxΩΩxS.....   43904.0(100.0)
       xxθθSS.....   142296.0(100.0) xxθΩSS.....   51744.0(100.0)
       xxΩΩSS.....   4704.0(100.0) xSθθSS.....   6776.0(100.0)
25   xSθΩSS.....   2464.0(100.0) xSΩΩSS.....   224.0(100.0)
       SSθθSS.....   121.0(100.0) SSθΩSS.....   44.0(100.0)
       SSΩΩSS.....   4.0(100.0)

142

Table 132: Relative efficiencies of various simple variegation codons

| | | Number of codons | | |
| | | 5 | 6 | 7 |
	vgCodon	#DNA/#AA [#DNA] (#AA)	#DNA/#AA [#DNA] (#AA)	#DNA/#AA [#DNA] (#AA)
10	NNK assuming stops vanish	8.95 $[2.86 \cdot 10^7]$ $(3.2 \cdot 10^6)$	13.86 $[8.87 \cdot 10^8]$ $(6.4 \cdot 10^7)$	21.49 $[2.75 \cdot 10^{10}]$ $(1.28 \cdot 10^9)$
15	NNT	1.38 $[1.05 \cdot 106]$ $(7.59 \cdot 10^5)$	1.47 $[1.68 \cdot 10^7]$ $(1.14 \cdot 10^7)$	1.57 $[2.68 \cdot 10^8]$ $(1.71 \cdot 10^8)$
20	NNG assuming stops vanish	2.04 $[7.59 \cdot 10^5]$ $(3.7 \cdot 10^5)$	2.36 $[1.14 \cdot 10^6]$ $(4.83 \cdot 10^6)$	2.72 $[1.71 \cdot 10^8]$ $(6.27 \cdot 10^7)$

Table 140.   **Effect f anti BPTI IgG on phage titer.**

Phage Strain	Input	+Anti-BPTI	+Anti-BPTI +Protein A(a)	Eluted Phage
M13MP18	100 (b)	98	92	$7 \cdot 10^{-4}$
BPTI.3	100	26	21	6
M13MB48 (c)	100	90	36	0.8
M13MB48 (d)	100	60	40	2.6

(a)   Protein A-agarose beads.

(b)   Percentage of input phage measured as plaque forming units

(c)   Batch number 3

(d)   Batch number 4

Table 141.   **Effect of anti-BPTI or protein A on phage titer.**

Strain	Input	No Addition	+Anti-BPTI	+Protein A (a)	+Anti-BPTI +Protein A
M13MP18	100(b)	107	105	72	65
M13MB48 (b)	100	92	7.10^{-3}	58	$\leq 10^{-4}$

(a)   Protein A-agarose beads

(b)   Percentage of input phage measured as plaque forming units

(c)   Batch number 5

144

Table 142 Effect of anti-BPTI and non-immune serum on phage titer

Strain	Input	+Anti-BPTI	+NRS (a)	+Anti-BPTI +Protein A (b)	+NRS +Protein A
M13MP18	100(c)	65	104	71	88
M13MB48(d)	100	30	125	13	121
M13MB48(e)	100	2	105	0.7	110

(a)   Purified IgG from normal rabbit serum.

(b)   Protein A-agarose beads.

(c)   Percentage of input phage measured as plaque forming units

(d)   Batch number 4

(e)   Batch number 5

Table 143. Loss in titer of display phage with anhydrotrypsin.

Strain	Anhydrotrypsin Beads		Streptavidin Beads	
	Start	Post Incubation	Start	Post Incubation
M13MP18	100 (a)	121	ND	ND
M13MB48	100	58	100	98
5AA Pool	100	44	100	93

(a) Plaque forming units expressed as a percentage of input.

Table 144. Binding of Display Phage t Anhydrotrypsin.

Experiment 1.

Strain	Eluted Phage (a)	Relative to M13MP18
M13MP18	0.2 (a)	1.0
BPTI-IIIMK	7.9	39.5
M13MB48	11.2	56.0

Experiment 2.

Strain	Eluted Phage (a)	Relative to M13mp18
M13mp18	0.3	1.0
BPTI-IIIMK	12.0	40.0
M13MB56	17.0	56.7

(a) Plaque forming units acid eluted from beads, expressed as a percentage of the input.

Table 145. Binding of Display Phage to Anhydrotrypsin or Trypsin.

Strain	Anhydrotrypsin Beads		Trypsin Beads	
	Eluted Phage (a)	Relative Binding (b)	Eluted Phage	Relative Binding
M13MP18	0.1	1	2.3×10^{-4}	1.0
BPTI-IIIMK	9.1	91	1.17	5x103
M13.3X7	25.0	250	1.4	6×10^3
M13.3X11	9.2	92	0.27	1.2×10^3

(a) Plaque forming units eluted from beads, expressed as a percentage of the input.
(b) Relative to the non-display phage, M13MP18.

146

Table 146. Binding of Display Phage to Trypsin or Human
Neutrophil Elastase.

Strain	Trypsin Beads			HNE Beads	
	Eluted Phage (a)	Relative Binding(b)		Eluted Phage	Relative Binding
M13MP18	5×10^{-4}	1		3×10^{-4}	1.0
BPTI-IIIMK	1.0	2000		5×10^{-3}	16.7
M13MB48	0.13	260		9×10^{-3}	30.0
M13.3X7	1.15	2300		1×10^{-3}	3.3
M13.3X11	0.8	1600		2×10^{-3}	6.7
BPTI3.CL (c)	1×10^{-3}	2		4.1	1.4×10^{4}

(a) Plaque forming units acid eluted from the beads,
expressed as a percentage of input.

(b) Relative to the non-display phage, M13MP18.

(c) BPTI-IIIMK (K15L MGNG)

147

Table 820: Streptavidin-Binding Phage

Putative Streptavidin

Name	Binding Peptide Seq.
DEV(F)	A E - P <u>C H P Q Y R L C</u> Q R P L K Q P P P P P P A E...
Dev(E)	A E - L <u>C H P Q F P R C</u> N L F R K V P P P P P P A E...
HPQ6	A E G P <u>C H P Q F P R C</u> Y I E G R I V - - - - - - E...

```
                1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6
- - - - C - - - - - - C - - - - - - - - - - - - - E
```

148

Table 827: Effect f DTT on biotinylated HRP binding to streptavidin agarose.

Conc. DTT (mM)	Biotin-HRP Color Development
0	++++
2	++++
10	+++
20	+
50	-

Table 828: Effect of DTT on HPQ6 display phage infectivity.

DTT (mM)	Relative Infectivity
0	1.00
2	0.95
10	1.00
20	1.08

149

50	1.23

Table 829: Effect of the presence of DTT on the binding of display phage to streptavidin agarose beads.

Inputs: MKTN 4.2×10^{11}, HPQ6 3.3×10^{11}.

Name	Concn DTT (μM)	Fraction Bound	Relative Binding
MKTN	0	4.8×10^{-6}	1.00
	2,000	5.4×10^{-6}	1.10
	10,000	5.2×10^{-6}	1.10
HPQ6	0	1.6×10^{-4}	1.00
	1	1.6×10^{-4}	1.00
	10	1.5×10^{-4}	0.90
	100	9.7×10^{-5}	0.60
	1000	1.6×10^{-5}	0.10
	2,000	1.0×10^{-5}	0.06
	5,000	8.8×10^{-6}	0.05

150

Table 832: Effect of preincubating HPQ6 display phage with Factor X$_a$ n binding to streptavidin beads.

Factor X$_a$	Titer after Treatment	Relative Titer	Fraction Bound	Relative Binding
-	3.3×10^{11}	1	1.4×10^{-4}	1
+	3.3×10^{11}	1	1.2×10^{-5}	8.5×10^{-2}

Table 833: FX$_a$ treatment of HPQ6 display phage following binding to streptavidin.

Factor X$_a$	Total Fraction Bound	Fraction Eluted by Treatment	% Removed by Treatment
-	7.6×10^{-3}	1.6×10^{-3}	14
+	6.4×10^{-3}	2.6×10^{-3}	40

Table 834: Removal of HPQ6 display phage from streptavidin by FX$_a$

Amount of FX$_a$ Added (units)	Time (hrs)	% Removed by Treatment
0	1	17
2.5	1	21
6.3	1	22
12.5	1	35
.0	2	18
2.5	2	53
6.3	2	54
12.5	2	52

CITATIONS

ASHM89:   Ashman, Matthews, and Frank (1989) Protein
Engineering 2(5): 387-91.

BANN81:   Banner, DW, C Nave, and DA Marvin, Nature (1981),
289:814-816.

BECK89c: Becker, S, E Atherton, and RD Gordon, Eur J Biochem,
(Oct 20 1989), 185(1)79-84.

BERG88:   Berg, JM, Proc Natl Acad Sci USA (1988), 85:99-102.

BETT88:   Better, M, CP Chang, RR Robinson, and AH Horwitz,
Science (1988), 240:1041-1043.

BOEK80:   Boeke, JD, M Russel, and P Model, J Mol Biol (1980),
144:103-116.

CHAN79:   Chang, CN, P Model, and G Blobel, Proc Natl Acad Sci
USA (1979), 76:1251-1255.

CHOU74:   Chou, PY, and GD Fasman, Biochemistry (1974),
13:(2)222-45.

COLM87:   Hemostasis and Thrombosis, Second Edition, Editors
Colman, Hirsh, Marder, and Salzman, Published by
Pippincott, Philadelphia, PA, 1987, ISBN 0-397-
50679-1.

CREI84:   Creighton, TE, Proteins: Structures and Molecular
Principles, W H Freeman & Co, New York, 1984.

CWIR90:   Cwirla, SE, EA Peters, RW Barrett, and WJ Dower,
Proc Natl Acad Sci USA, (August 1990), 87:6378-6382.

DELA88:   de la Cruz, VF, AA Lal and TF McCutchan,
J Biol Chem, (1988), 263(9)4318-22.

DEVL90:   Devlin, JJ, LC Panganiban, and PE Devlin, Science,
(27 July 1990), 249:404-406.

DICK83:   Dickerson, RE, and I Geis, Hemoglobin: Structure,
Function, Evolution, and Pathology, The
Bejamin/Cummings Publishing Co, Menlo Park, CA,
1983.

DULB86:   Dulbecco, R, US Patent 4,593,002, June 3, 1986.

152

GAUS87:   Gauss, P, KB Krassa, DS McPheeters, MA Nelson, and L
           Gold, Proc Natl Acad Sci USA (1987), 84:8515-19.

GIBS88:   Gibson, TJ, JPM Postma, RS Brown, and P Argos,
           Protein Engineering (1988), 2(3)209-218.

GUAN91:   Guan, KL and Dixon JE, Anal. Biochem, (1991), 192:
           262-67.

HARD90:   Hard, T, E Kellenbach, R Boelens, BA Maler, K
           Dahlman, LP Freedman, J Carlstedt-Duke, KR Yamamoto,
           J-A Gustafsson, and R Kaptein, Science (13 July
           1990), 249:157-60.

HORV89:   Horvat, S, B Grgas, N Raos, and VI Simeon, Int J
           Peptide Protein Res (1989), 34:346-51.

INOU82:   Inouye, H, W Barnes, and J Beckwith, J Bacteriol
           (1982), 149(2)434-439.

ITOK79:   Ito, K, G Mandel, and W Wickner, Proc Natl Acad Sci
           USA (1979), 76:1199-1203.

JANA89:   Janatova, J, KBM Reid, and AC Willis, Biochem
           (1989), 28:4754-61.

JANI85:   Janin, J, and C Chothia, Methods in Enzymology
           (1985), 115(28)420-430.

KAPL78:   Kaplan, DA, L Greenfield, and G Wilcox, in The
           Single-Stranded DNA Phages, Denhardt, DT, D
           Dressler, and DS Ray editors, Cold Spring Harbor
           Laboratory, 1978., p461-467.

KATZ90:   Katz, B, and AA Kossiakoff, Proteins, Struct, Funct,
           and Genet (1990), 7:343-57.

KIMH89:   Kim, et al., Protein Engineering (1989), 2(1): 379-
           86.

KISH85:   Kishore, R, and P Balaram, Biopolymers (1985),
           24:2041-43.

KUHN85a: Kuhn, A, and W Wickner, J Biol Chem (1985),
            260:15914-15918.

KUHN85b: Kuhn, A, and W Wickner, J Biol Chem (1985),
            260:15907-15913.

KUHN87:   Kuhn, A, Science (1987), 238:1413-1415.

LISS85:   Liss, LR, BL Johnson, and DB Oliver, J Bacteriol
            (1985), 164(2)925-8.

LOPE85a: Lopez, J, and RE Webster, J Bacteriol (1985),
            163(3)1270-4.

LUIT85:   Luiten, RGM, DG Putterman, JGG Schoenmakers, RNH
            Konings, and LA Day, J Virology, (1985), 56(1)268-
            276.

LUIT87:   Luiten, RGM, RIL Eggen, JGG Schoenmakers, and RNH
            Konings, DNA (1987), 6(2)129-37.

MAKO80:   Makowski, L, DLD Caspar, and DA Marvin, J Mol Biol
            (1980), 140:149-181.

MANI82:   Maniatis, T, EF Fritsch, and J Sambrook, Molecular
            Cloning, Cold Spring Harbor Laboratory, 1982.

MARK86:   Marks, CB, M Vasser, P Ng, W Henzel, and S Anderson,
            J Biol Chem (1986), 261:7115-7118.

MARV78:   Marvin, DA, in The Single-Stranded DNA Phages,
            Denhardt, DT, D Dressler, and DS Ray editors, Cold
            Spring Harbor Laboratory, 1978., p583-603.

MATS89:   Matsumura, M, WJ Becktel, M Levitt, and BW Matthews,
            Proc Natl Acad Sci USA (1989), 86:6562-6.

MCCA90:   McCafferty, J, AD Griffiths, G Winter, and DJ
            Chiswell, Nature, (6 Dec 1990), 348:552-4.

MESS77:   Messing, J, B Gronenborn, B Muller-Hill, and
            PH Hofschneider, Proc Natl Acad Sci USA (1977),
            74:3642-6.

MESS78:   Messing, J, and B Gronenborn, in The Single-Stranded

DNA Phages, Denhardt, DT, D Dressler, and DS Ray
editors, Cold Spring Harbor Laboratory, 1978.,p449-
453.

NICH88:   Nicholson, H, WJ Becktel, and BW MAtthews, Nature
           (1988), 336:651-56.

NISH82:   Nishiuchi, Y, and S Sakakibara, FEBS Lett (1982),
           148:260-2.

NISH86:   Nishiuchi, Y, K Kumagaye, Y Noda, TX Watanabe, and S
           Sakakibara, Biopolymers, (1986), 25:S61-8.

OHKA81:   Ohkawa, I, and RE Webster, J Biol Chem (1981),
           256:9951-9958.

OLIV90a: Olivera, BM, J Rivier, C Clark, CA Ramilo, GP
           Corpuz, FC Abogadie, EE Mena, SR Woodward, DR
           Hillyard, LJ Cruz, Science, (20 July 1990), 249:257-
           263.

PABO79:   Pabo, CO, RT Sauer, JM Sturtevant, and M Ptashne,
           Proc Natl Acad Sci USA (1979), 76:1608-1612.

PARM88:   Parmley, SF, and GP Smith, Gene (1988), 73:305-318.

PERR84:   Perry, LJ, and R Wetzel, Science (1984), 226:555-7.

PERR86:   Perry, LJ, and R Wetzel, Biochem (1986), 25:733-39.

POTE83:   Poteete, AR, J Mol Biol (1983), 171:401-418.

RASC86:   Rasched, I, and E Oberer, Microbiol Rev (1986)
           50:401-427.

RASH84:   Rashin, A, Biochemistry (1984), 23:5518.

ROSE85:   Rose, GD, Methods in Enzymololgy (1985), 115(29)430-
           440.

RUSS81:   Russel, M, and P Model, Proc Natl Acad Sci USA
           (1981), 78:1717-1721.

RUSS82:   Russel and Model, Cell (1982), 28(1): 177-84.

SALI88:   Sali, D, M Bycroft, and AR Fersht, Nature (1988),
           335:740-3.

SAMB89:   Sambrook, J, EF Fritsch, and T Maniatis, Molecular
           Cloning, A Laboratory Manual, Second Edition, Cold
           Spring Harbor Laboratory, 1989.

SAUE86:   Sauer, RT, K Hehir, RS Stearman, MA Weiss,
            A Jeitler-Nilsson, EG Suchanek, and CO Pabo, Biochem
            (1986), 25:5992-98.

SCHA78:   Schaller, H, E Beck, and M Takanami, in The Single-
            Stranded DNA Phages, Denhardt, D.T., D. Dressler,
            and D.S. Ray editors, Cold Spring Harbor Laboratory,
            1978., p139-163.

SCHN86:   Schnabel, E, W Schroeder, and G Reinhardt, Biol Chem
            Hoppe-Seyler (1986), 367:1167-76.

SCOT87a: Scott, MJ, CS Huckaby, I Kato, WJ Kohr, M Laskowski
            Jr., M-J Tsai and BW O'Malley, J Biol Chem (1987),
            262(12)5899-5907.

SCOT90:   Scott, JK, and GP Smith, Science, (27 July 1990),
            249:386-390.

SMIT85:   Smith GP, Science (1985), 228:1315-1317.

SUMM91:   Summers, J. Cell Biochem. (1991), 45(1): 41-8.

VITA84:   Vita, C, D Dalzoppo, and A Fontana, Biochemistry
            (1984), 23:5512-5519.

WEBS78:   Webster, RE, and JS Cashman, in The Single-Stranded
            DNA Phages, Denhardt, DT, D Dressler, and DS Ray
            editors, Cold Spring Harbor Laboratory, 1978., p557-
            569.

WELL86:   Wells, JA, and DB Powers, J Biol Chem (1986),
            261:6564-70.

YANI85:   Yanisch-Perron, C, J Vieira, and J Messing, Gene,
            (1985), 33:103-119.

ZAFA88:   Zafaralla, GC, C Ramilo, WR Gray, R Karlstrom, BM
            Olivera, and LJ Cruz, Biochemistry, (1988),
            27(18)7102-5.

ZIMM82:   Zimmermann, R, C Watts, and W Wickner, J Biol Chem
            (1982), 257:6529-6536.

156

CLAIMS

        1. In a process for developing novel epitopes or binding proteins with a desired binding activity against a particular target material which comprises providing a
5 library of phage which each displays on its surface, as a result of expression of a first phage gene, one or more copies of a particular chimeric coat protein, each chimeric coat protein comprising a potential epitope, or a potential binding domain which is a mutant of a known protein domain
10 foreign to said phage, said library collectively displaying a plurality of potential epitopes or binding domains, contacting said library of phage with the target material, and separating the phage on the basis of their affinity for the target material, the improvement wherein said chimeric
15 coat protein further comprises a linker peptide which is specifically cleavable by said site-specific protease.

        2. The method of claim 1 wherein the site-specific protease is Factor Xa, Factor XIa, kallikrein, thrombin, Factor XIIa, collagenase or enterokinase.

20   3. The method of claim 1 wherein, after said library of phage is contacted with said target material, (1) low affinity phage are removed, (2) high affinity phage still bound to said target material are released by cleavage of said chimeric coat protein at said linker by
25 means of a site-specific protease, and the released high affinity phage are recovered.

        4. In a process for developing novel binding peptides or proteins with a desired binding activity against a particular target material which comprises
30 providing a library of phage which each displays on its surface, as a result of expression of a first phage gene, one or more copies of a particular chimeric coat protein, each chimeric coat protein comprising a mutant of a known protein domain foreign to said phage, said library

collectively displaying a plurality of potential binding
domains, contacting said library of phage with the target
material, and separating the phage on the basis of their
affinity for the target material, the improvement wherein
5   said said potential binding domain has at least one
intrachain covalent crosslink between a first amino acid
position and a second amino acid position thereof, the
amino acids at said first and second positions being
invariant in all of the chimeric proteins displayed by said
10   library, and where low affinity phage are removed from said
target material first, and then high affinity phage are
released or rendered more readily eluted from the target
material by treating the phage with a reagent which cleaves
the crosslink, preferably a reagent which does not kill the
15   phage.
             5. The method of any of claims 1-4 wherein the
domain is a mini-protein of less than sixty amino acids,
more preferably a micro-protein of less than forty amino
acids.
20   6. The method of claim 4 wherein the crosslink
is a disulfide bond and the amino acids at said first and
second positions are cysteines.
             7. The method of claim 6 wherein the reagent is
dithiothreitol.
25   8. The method of any of claims 1-7 wherein the
phage further comprises a second phage gene encoding the
cognate wild-type coat protein of the phage.
             9. In a process for developing novel epitopes
with a desired binding activity against a particular target
30   material which comprises providing a library of phage which
each displays on its surface, as a result of expression of
a first phage gene, one or more copies of a particular
chimeric coat protein, each chimeric coat protein
comprising a potential epitope, said library collectively

displaying a plurality of potential epitopes, contacting
said library of phage with the target material, and
separating the phage on the basis of their affinity for the
target material, the improvement wherein the phage further

5   comprises a second phage gene encoding the cognate wild-
type coat protein of the phage.

        10. In a process for developing novel epitopes
or binding proteins with a desired binding activity against
a particular target material which comprises providing a

10   library of phage which each displays on its surface, as a
result of expression of a first phage gene, one or more
copies of a particular chimeric coat protein, each chimeric
coat protein comprising a potential epitope, or a potential
binding domain which is a mutant of a known protein domain

15   foreign to said phage, said library collectively displaying
a plurality of potential epitopes or binding domains,
contacting said library of phage with the target material,
and separating the phage on the basis of their affinity for
the target material, the improvement wherein the chimeric

20   coat protein includes only an assemblable fragment of a
coat protein of said phage, and not that portion of the
coat protein which is responsible for pilus binding, and
the phage also comprises a second phage gene encoding the
cognate native coat protein of the phage.

25   11. In a process for developing novel epitopes
with a desired binding activity against a particular target
material which comprises providing a library of phage which
each displays on its surface, as a result of expression of
a first phage gene, one or more copies of a particular

30   chimeric coat protein, each chimeric coat protein
comprising a potential epitope, said library collectively
displaying a plurality of potential epitopes, contacting
said library of phage with the target material, and
separating the phage on the basis of their affinity for the

target material, the improvement wherein the cognate wild-
type coat protein of the phage is the major coat protein of
the phage.

      12. The method of claims 8-11 wherein the
5   initiation codon of the second phage gene is a leucine.

      13. The method of any of claims 1-12 wherein the
first phage gene further comprises a cytoplasmic secretion
signal sequence which codes for a signal peptide which
directs the immediate expression product to the inner
10   membrane of the bacterial host cell infected by said phage,
where it is processed to remove said signal peptide,
yielding a mature chimeric coat protein comprising the
potential binding domain and at least a portion of a
geneVIII-like protein of the phage, said chimeric protein
15   being assembled with wild-type coat protein into the phage
coat, wherein the secretion signal is encoded by a signal
sequence selected from the group consisting of the signal
sequences of the phoA, bla and geneIII genes.

      14. In a process for developing novel epitopes
20   or binding proteins with a desired binding activity against
a particular target material which comprises providing a
library of phage which each displays on its surface, as a
result of expression of a first phage gene, one or more
copies of a particular chimeric coat protein, each chimeric
25   coat protein comprising a potential epitope, or a potential
binding domain which is a mutant of a known protein domain
foreign to said phage, said library collectively displaying
a plurality of potential epitopes or binding domains,
contacting said library of phage with the target material,
30   and separating the phage on the basis of their affinity for
the target material, the improvement wherein the phage also
comprises a second phage gene encoding the cognate native
coat protein of the phage, and the initiation codon of the
second phage gene is a Leucine codon.

15. The method of any of claims 1-14 wherein the
differentiation among said plurality of different potential
binding domains occurs through the at least partially
random variation of one or more predetermined amino acid
5   positions of said known domain to randomly obtain at each
said position an amino acid belonging to a predetermined
set of two or more amino acids, the amino acids of said set
occurring at said position in predetermined expected
proportions.

10   16. The method of claim 15 wherein the differentiation
among said potential binding domains of said library is limited to
no more than about 20 predetermined amino acid residues of said
sequence.

17. The method of claim 15 wherein, for each set, the
15 ratio of the probability of occurrence of the most favored amino
acid to that for the least favored amino acid is less than aboout
2.6.

18. The method of any of claims 1-17 wherein, for any
20 potentially encoded potential binding domain, the probability that
it will be displayed by at least one package in said population is
at least 50%, more preferably at least 90%.

19. The method of any of claims 1-18 wherein said
population is characterized by the display of at least 10^5
25 different potential binding domains.

20. The method of any of claims 1-19 wherein the
initially chosen parental potential binding domain is selected from
the group consisting of (a) binding domains of bovine pancreatic
trypsin inhibitor, crambin, Cucurbita maxima trypsin inhibitor III,
30 a heat-stable enterotoxin of Excherichia coli, an alpha-, mu- or
omega-conotoxin, apamin, charybdotoxin, secretory leukocyte
protease inhibitor, cystatin, eglin, barley protease inhibitor,
ovomucoid, T4 lysozyme, hen egg white lysozyme, ribonuclease,
azurin, tumor necrosis factor, and CD4, and (b) domains at least

substantially homologous with any of the foregoing domains which
have a melting point of at least 50°C.

     21. In a process for developing novel epitopes with a
desired affinity for a particular binding protein target material
5 which comprises providing a library of phage which each displays on
its surface, as a result of expression of a first phage gene, one
or more copies of a particular chimeric coat protein, each chimeric
coat protein comprising a potential epitope, or a potential binding
domain which is a mutant of a known protein domain foreign to said
10 phage, said library collectively displaying a plurality of
potential epitopes or binding domains, contacting said library of
phage with the target material, and separating the phage on the
basis of their affinity for the binding protein target material,
the improvement wherein the differentiation among said plurality
15 of different potential binding domains occurs through the at least
partially random variation of one or more predetermined amino acid
positions of said known domain to randomly obtain at each said
position an amino acid belonging to a predetermined set of two or
more amino acids, the amino acids of said set occurring at said
20 position in predetermined expected proportions, and in
substantially all sets the ratio of the frequency of occurrence of
the most favored amino acid to that for the least favored amino
acid is less than 2.6.

     22. The method of claim 21 in which at least one
25 variable amino acid position is encoded by a simply variegated
codon selected from the group consisting of NNT, NNG, RNG, RMG,
VNT, RRS, and SNT.

     23. The method of claim 21 wherein none of the variable
amino acid positions is encoded by a simply variegated codon
30 selected from the group consisting of NNN, NNK and NNS.

     24. The method of claim 21 in which at least one
variable amino acid position is encoded by a complexly variegated
codon.

162

25. A library of display phage which each displays on
its surface, as a result of expression of a first phage gene, one
or more copies of a particular chimeric coat protein, each chimeric
coat protein comprising a potential epitope, or a potential binding
5 domain which is a mutant of a known protein domain foreign to said
phage, said library collectively displaying a plurality of
potential epitopes or binding domains, contacting said library of
phage with the target material, and separating the phage on the
basis of their affinity for the binding protein target material,
10 wherein the differentiation among said plurality of different
potential binding domains occurs through the at least partially
random variation of one or more predetermined amino acid positions
of said known domain to randomly obtain at each said position an
amino acid belonging to a predetermined set of two or more amino
15 acids, the amino acids of said set occurring at said position in
predetermined expected proportions, and in substantially all sets
the ratio of the frequency of occurrence of the most favored amino
acid to that for the least favored amino acid is less than 2.6.

.26. A library of display phage which each displays on
20 its surface, as a result of expression of a first phage gene, one
or more copies of a particular chimeric coat protein, each chimeric
coat protein comprising a potential epitope, or a potential binding
domain which is a mutant of a known protein domain foreign to said
phage, said library collectively displaying a plurality of
25 potential epitopes or binding domains, contacting said library of
phage with the target material, and separating the phage on the
basis of their affinity for the target material, wherein said
chimeric coat protein further comprises a linker peptide which is
specifically cleavable by said site-specific protease.

30

1/2

FIG. 1.
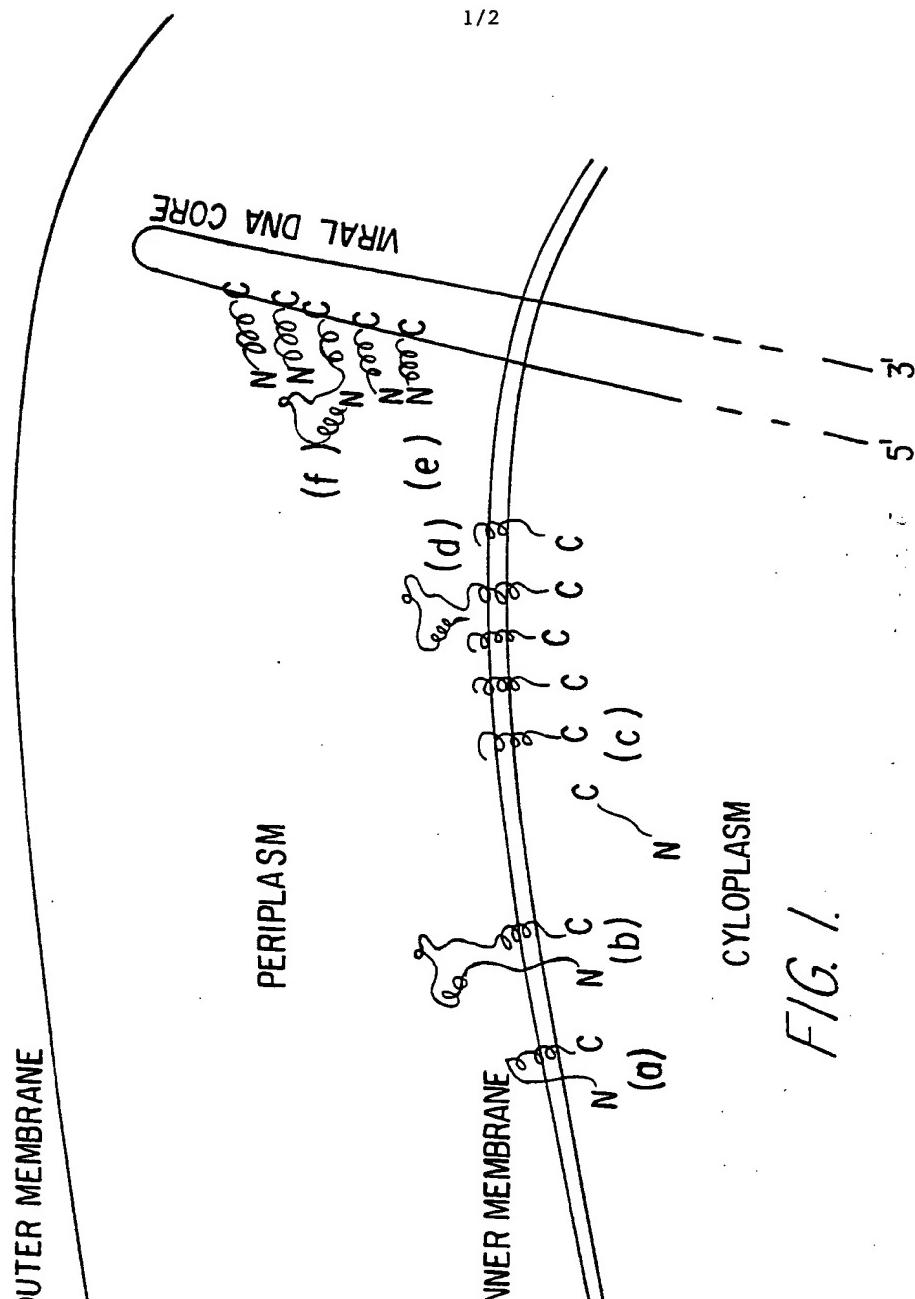
FIG 2.
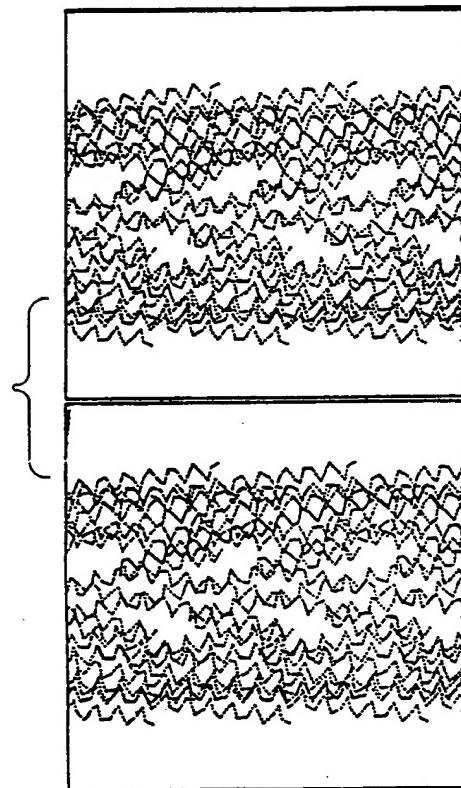
I. CLASSIFICATION OF SUBJECT MATTER   (If several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/10;   C12N15/62;   C12N15/70;   C12N15/12
         C12N15/66;   C12P21/02;   C07K13/00

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N ; C12P ; C07K

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category[o]	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
X	WO,A,9 002 809 (PROTEIN ENGINEERING CORPORATION) 22 March 1990 cited in the application	4,8-9, 11, 15-22, 24-25
Y	see page 97 - page 102 see page 108 - page 109	1-3,5-7, 13,26
	see page 122 - page 153; claims	

Y	NATURE. vol. 309, 28 June 1984, LONDON GB pages 810 - 812; NAGAI, K. & THOGERSEN, H.C.: 'Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in Escherichia coli' see the whole document	1-3,26

	-/--	

° Special categories of cited documents:[10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
14 JULY 1992	2 4. 07. 92

International Searching Authority	Signature of Authorized Officer
EUR PEAN PATENT OFFICE	ANDRES S.M.

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category °	Citation of Document, with indication, where appropriate, of the relevant passage	Relevant to Claim No.
Y	JOURNAL OF BIOLOGICAL CHEMISTRY. vol. 261, no. 16, 5 June 1986, BALTIMORE US pages 7115 - 7118; MARKS, C.B. ET AL.: 'Production of native, correctly folded bovine pancreatic trypsin inhibitor by Escherichia coli' cited in the application see page 7115, right column, paragraph 3 ---	13
Y	SCIENCE. vol. 249, 20 July 1990, LANCASTER, PA US pages 257 - 263; OLIVERA, B.M. ET AL.: 'Diversity of Conus neuropeptides' cited in the application see the whole document ---	5-7
A	SCIENCE. vol. 249, 27 July 1990, LANCASTER, PA US pages 404 - 406; DEVLIN, J.J. ET AL.: 'Random peptide libraries: a source of specific protein binding molecules' cited in the application see the whole document ---	
A	WO,A,8 902 461 (SCHERING AKTIENGESELLSCHAFT) 23 March 1989 see abstract; claims; figures ---	1-2,13, 26
P,X	WO,A,9 201 047 (CAMBRIDGE ANTIBODY TECHNOLOGY LIMITED & MEDICAL RESEARCH COUNCIL) 23 January 1992 see page 21, line 7 - line 40 see page 24, line 49 - page 25, line 4 see page 31, line 3 - line 48; examples 24,47 ---	1-4,6-7, 26
P,X	GENE. vol. 109, no. 1, 20 December 1991, AMSTERDAM NL pages 13 - 19; MARKLAND, W. ET AL.: 'Design, construction and function of a multicopy display vector using fusions to the major coat protein of bacteriophage M13' see the whole document ---	9,11-15, 20

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information. 14/07/92

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9002809	22-03-90	AU-A-	4308689	02-04-90
		EP-A-	0436597	17-07-91
		JP-T-	4502700	21-05-92
WO-A-8902461	23-03-89	DE-A-	3731874	30-03-89
		EP-A-	0393039	24-10-90
		JP-T-	3500241	24-01-91
WO-A-9201047	23-01-92	AU-A-	8221691	04-02-92

THIS PAGE BLANK (USPTO)